Development and use of engineered peptide deformylase in chemoenzymatic peptide synthesis

Di Toma, Claudia

Development and Use of Engineered Peptide Deformylases in Chemoenzymatic Peptide Synthesis

Painting on cover page by G. Centazzo – Acque del Tagliamento – Olio su tela 70X80 cm

Printed by: Grafimedia Facilitair Bedrijf RUG

The research presented in this thesis was carried out at DSM Innovative Synthesis B.V. in Geleen.

RIJKSUNIVERSITEIT GRONINGEN

Development and Use of Engineered Peptide Deformylases in Chemoenzymatic peptide Synthesis

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag 5 oktober 2012
om 12.45 uur

door

Claudia Di Toma

geboren op 23 oktober 1979

te Gemona del Friuli, Italië

Promotor:	Prof. dr. D.B. Janssen
Copromotor:	Dr. T. Sonke
Beoordelingscommissie:	Prof. dr. L. Dijkhuizen Prof. dr. R. Bovenberg Prof. dr. U. Hanefeld

ISBN:978-90-367-5796-8
ISBN:978-90-367-5796-6 (Electronic version)

Contents

Chapter 1

General introduction

General introduction

Peptides (from the Greek *πεπτίδια*, "small digestible") are compounds consisting of chains of amino acids that are linked in a defined order by amide bonds. Peptides with a molecular mass exceeding 6,000 Da are usually called proteins, many of which act as enzymes or play a structural role. Smaller peptides normally have no catalytic activity, but they often are involved in the regulation of physiological processes in living cells and organisms, and therefore may be of high medical relevance. Besides that, many peptides are important for synthetic chemistry, biotechnology, and food technology. For example, because of their special taste or antimicrobial activity, peptides are widely recognized as important food components and they may be used as food additives [45].

The discovery of new peptides, the elucidation of their physiological role and bioactivity, the development of new delivery strategies, as well as the design of better methods for their synthesis and purification have contributed to an expanding interest in exploring new practical applications of peptides. This growing interest, in turn, has stimulated the demand for peptides in large amounts, and thus for cheaper, more efficient, and greener methods for peptide synthesis than those traditionally used under laboratory conditions.

Therapeutic peptides

Potential of peptides as therapeutic agents. Many peptides display a bioactivity that has important therapeutic potential, for example in blood pressure control, as growth-stimulating hormone, or as agents that regulate important processes like digestion and reproduction. Examples of such peptides are corticotropin (stimulates production of adrenocorticotropic hormone (ACTH), which influences steroid metabolism), growth hormone, arginine- and lysine-vasopressins (used for controlling pressor and antidiuretic activitities), insulin (glucose metabolism regulator, indicated for treatment of type 1 or 2 diabetes mellitus) and glucagon-like peptide-1 (also used in diabetes therapy), defensin (antimicrobial agent that functions in the immune system by permeabilizing bacterial membranes) and gastrin-releasing peptides (control functions of the gastrointestinal system)[114]. Another particularly promising area of biomedical application is the prevention of infections by inhibition of host-pathogen interactions

through mimicking binding sites of the pathogen or host proteins [28]. Also the classical β-lactam antibiotics can be regarded as peptides or peptide-derived compounds. See Table 1 for a number of examples of bioactive peptides and their activity and application.

Limitations & obstacles. For many years, the major obstacle to the success of using peptides as pharmaceuticals was their lack of oral availability. As a result, relatively few peptides reached the market as approved drugs. Consequently, several major pharmaceutical companies abandoned their research efforts in this area, in favour of investigations on small-molecule mimics of peptides or proteins that can be considered as lead compounds [114]. In recent years, however, advances in formulation and the development of novel drug delivery systems have revitalized the field, leading to several highly promising peptide drugs {Futaki, Suzuki, et al. 2001 355 /id}[3,34]. For example, peptide conjugates with doxorubicin covalently linked to neuropeptide Y have been described for tumor-specific chemotherapy, and peptide drugs unable to pass the blood-brain barrier could be delivered using a chimeric strategy in which non-transportable drugs were coupled with blood-brain barrier transport vectors like cationized albumin or antibodies that undergo receptor-mediated transcytosis. This way pharmacologic effects in the central nervous system were achieved [11;66].

Besides difficulties with delivery, the stability of peptides in the body may be an issue. Despite the high affinity of peptides to their specific target, the use of peptides has been limited due to their susceptibility to proteolytic enzymes that are widely distributed in body tissues. To make peptides pharmacologically more useful, it is necessary to increase their serum half-life by enhancing resistance to proteolysis. The pathways of proteolysis can be difficult to determine but usually an inspection of a peptide sequence provides clues about the possible sites of proteolytic degradation. Peptidases can be divided in exopeptidases, which cleave peptide bonds at the termini of the peptide chain, and endopeptidases, which cleave internal peptide bonds. Endopeptidases have usually specificity for a particular set of amino acids, whereas exopeptidases can be divided in aminopeptidases, cleaving off the *N*-terminal amino acid, and carboxypeptidases, which cleave the *C*-terminal amino acid from a peptide chain. Different strategies can be applied to stabilize a peptide against proteolysis by peptidases. For example acetylation of the amino terminus can block aminopeptidases, and endoprotease cleavage can be suppressed by modification of amino acid side chains (e.g. glycosylation), side chain cross-linking (to enhance α-helix conformational

stability) or inversion of chirality [96;108]. The peptide bond itself can be replaced by hydrolytically inert isosteric structures like *trans* olefins or by substitution of α-hydrogens or amide hydrogens by methyl groups [6]. Despite their efficacy with regard to increasing peptide stability, it has to be kept in mind that all these methods involve some modification of the peptide and therefore can also influence the biological activity of the molecule.

Sources of therapeutic peptides. Until the late 1970s many naturally occurring peptides and proteins for therapeutic use were isolated from human tissue or urine, including insulin, growth hormone, and coagulation factor VIII [114]. Other natural sources such as bacterial and fungal cultures, plant biomass, or animal tissue were also used [27].

As with most drug classes, non-natural therapeutic peptides with new properties, were developed through the confluence of improved tools for chemical synthesis and analysis, as well as a growing understanding of their biological action. Peptide leads with new activities that were traditionally obtained in low amounts by modification of natural peptides are nowadays rapidly prepared in the form of large libraries by chemical (combinatorial) synthesis. Computer docking using receptor structures (in silico screening), genome sequencing, and research in cell biology have identified many new potential targets. For activity screening, high-throughput methods and more sensitive assays have become available [107].

Market. The general property that makes peptides attractive therapeutic agents and potential blockbuster drugs are their high target specificities [23;34;111]. Today, more than 60 different peptides are on the market, around 270 are in clinical trials [128;131]. This field is rapidly expanding, and the annual market growth of pharmaceutical peptides (7-15% per year) clearly exceeds the average growth of pharmaceuticals (~5% per year). This increasing demand is especially due to the growing need for synthetic peptides that can be used in therapeutic research and in clinical development. Furthermore, the availability of peptide libraries for drug screening, epitope mapping and structure-activity studies increases the number of hits that require further exploration. The growing interest includes the use of peptides as carriers for the delivery of exogenous proteins into living cells.

Table 1. Examples of therapeutic peptides and their activity or application.

Peptide	Activity	Effect or application
Oxytocin and analogues (9 aa)	peripheral (hormonal) actions/ action on oxytocin receptors in the brain and spinal cord	labor induction
Vasopressin analogues (9 aa)	regulation of body's water retention increasing arterial blood pressure	antidiuretic/ vasoconstriction/ esophageal varices/ septic shock
Insulin (21 + 30 aa)	regulation of glucose intake/regulation of DNA replication and protein synthesis	diabetes/ insulinoma/ metabolic syndrome
Glucagon (29 aa)	regulation of carbohydrate metabolism/ stimulation of insulin release	hypoglycemia
Secretin (27 aa)	control of duodenal secretions and water homeostasis/ stimulation of insulin release/inhibition of gastrin release	pancreatic functioning test
Calcitonin (32 aa)	regulation of bone minerals metabolism/regulation of feeding and appetite, calcium homeostasis	postmenopausal osteoporosis/ hypercalcemia/ paget's disease/ bone metastases
Luteinizing hormone-releasing hormone (LH-RH) and analogues (10 aa)	stimulus for ovulation and corpus luteum maturation/ regulation of testosterone production	hypothalamic hypogonadism/ kallmann syndrome/ infertility therapy
Parathyroid hormone (PTH)(84 aa)	regulation of serum calcium levels/ stimulus vitamin D synthesis	hypoparathyroidism
Corticoliberin- releasing hormone (CRH)(41 aa)	trigger for parturition/ stimulation of ACTH production	generalized anxiety disorders/depression
Adrenocorticotropic hormone (ACTH)(39 aa)	regulation of steroidogenesis/ regulation of cholesterol transport in mitochondria/ stimulation of lipoprotein uptake	adrenal insufficiency/ ACTH stimulation test
Growth hormone-releasing hormone (44 aa)	stimulation of growth hormone production	dwarfism
Somatostain and analogues (14 aa)	endocrine regulation, cell proliferation	acromegaly/ neuroendocrine tumors/ acute variceal bleeding
Thyroliberin (TRH)(3 aa)	stimulation of pituitary gland	diagnostic tests of thyroid disorders/ acromegaly
Thymosin-α_1 (28 aa)	regulation of unpolymerized actin/ wound healing	immunopotentiation, anti-viral agent
Thymopentin (TP-5) (5 aa)	Immunostimulant	immune system stimulation
Cyclosporin (11 aa)	Immunosuppressor	post-allogeneic organ transplant treatment
Integrilin (8 aa)	blocks platelet glycoprotein receptor	acute cardiac ischemic events/ angina/ myocardial infarction

Data adopted from Andersson et al. [67] and papers cited in the text.

Structures of bioactive peptides. The types of bioactive peptides that can be used in medical applications are highly diverse, ranging from simple dipeptides to more complex linear or cyclic structures, which are often modified through glycosylation, phosphorylation or acylation of amino acid residues. The structure presents both key residues for functionality as well as a stable fold that serves as scaffold. The diversity and complexity of the modifications that are observed in bioactive peptides clearly reflect the versatility of functions in cell physiology and regulation.

As mentioned above, several artificial modifications can alter the properties of peptides in such a way that degradation is prevented and stability increased. Thus, modification at the *C*- or *N*-terminus may be used to block degradation by exo-peptidases [123]. Examples are *N*-terminal acetylation or glycosylation and *C*-terminal amidation. The use of unnatural amino acids as building blocks or the introduction of disulfide bonds may prevent cleavage of the peptide chain by endoproteases [65]. Polyethylene glycol (PEG) can also be attached at the N or C termini of peptides to increase their overall stability *in vivo* and prevent renal clearance [107]. Hematide™ is a synthetic peptide analog of erythropoietin, substituted with PEG to extent its *in vivo* half-life [122;130]. Other strategies to extend plasma residence time are chemical conjugation of peptides to albumin [4], to Fabs [90] or to scaffolds such as fibronectin or ankyrin repeats [13].

Peptides as food ingredients

Bioactive food peptides. Oligopeptides are also increasingly important as health-enhancing food ingredients. Furthermore, the importance of peptides in the sensory perception of food has been recognized for a long time [12]. Many biologically active peptides in food exert — beyond their nutritional value — a physiological, hormone-like effect in humans. In many diets, bovine milk and dairy products including cheese are the largest source of bioactive peptides. In addition, bioactive peptides and proteins are present in eggs, gelatin, meat and fish of various kinds, as well as in many plants (e.g. wheat, soy, rice)[86]. Most bioactive peptides that occur as food ingredients are formed by processing of larger polypeptides. They may be inactive within the sequence of their parent protein and can be released through hydrolysis by digestive enzymes (e.g. trypsin)[138], by proteases from microorganisms, or through the action of proteolytic enzymes from plants. Activation may occur either during gastrointestinal digestion or already during food processing (e.g. during cheese ripening and milk

fermentation) [61]. Food peptides that contribute to taste or that show activities that can be related to health issues usually contain 2–20 amino acid residues per molecule, but in some cases they may consist of more than 20 amino acids.

Bioactivity and health effects. After digestion, bioactive peptides from food can either be absorbed in the gastrointestinal tract to enter the blood circulation in intact form and then exert systemic effects, or they may produce local effects in the gastrointestinal tract [110]. Absorbed peptides can exhibit diverse activities. Potent ACE (angiotensin converting enzyme)-inhibiting peptides that are present in dairy products, fish and soy can cause reduction of blood pressure [29] and may improve blood lipid profiles [124]. Another example is the antioxidant activity of peptides released from casein during hydrolysis by digestive enzymes [106], which exhibit similar antioxidant activity as vitamin C and E, polyphenols and carotenoids released from vegetarian food sources. Cyto- or immunomodulatory effects such as modulation of lymphocyte proliferation can be stimulated by casein [87], lactoferrin [104], lactoperoxidase [135] and milk immunoglobulin G, while antithrombotic activities are reported for peptides derived by hydrolysis from κ-casein [86]. Many studies have been devoted to the effects of peptides derived from milk proteins. Caseinophosphopeptides (CPP) were reported to have an influence on the regulation of digestive enzymes and modulation of nutrient absorption via an opioid activity [61] but they can also enhance the mucosal immunity [73]..

Because of their health-enhancing potential and safety profiles, bioactive peptides may be used as additives or their level may be enriched by modification of a manufacturing process in functional foods and nutraceuticals. The results are attractive products for health-conscious people and persons with specific demands [29]. Consequently, a significant research effort has been aimed at characterizing and controlling protein hydrolysis technology to obtain bioactive peptides and as a result of these studies, numerous new products are now on the market or under development by food companies [46]. One example is the sour milk product Calpis™, available on the market, which contains antihypertensive tripeptides proven to have beneficial effects on blood pressure in human trials [112]. Another example is Nisaplin®, a commercial product having nisin as active component. Nisin is an antibacterial peptide that exerts a broad activity against Gram-positive bacteria by inhibiting peptidoglycan biosynthesis and by forming pores in the bacterial membrane. Nisaplin® was approved as antimicrobial for use in food [32].

The challenge for food technology is the development of foods and nutraceuticals in which the active compound is able to resist degradation during digestion and can be absorbed, attaining appropriate concentrations in serum. Moreover, bioactive ingredients need to possess a good shelf stability and the undesired bitter taste that is characteristic of many peptide hydrolysates has to be avoided [86].

Methods for peptide synthesis

Need and sources of peptides

The growing interest in the use of peptides requires cost-efficient methodologies for their production, isolation, and analysis. Three main approaches to peptide synthesis are currently available, *i.e.* solid-phase or solution-phase chemical peptide synthesis, chemo-enzymatic peptide synthesis, and fermentative peptide production. Advantages and drawbacks of the different methods are presented in the following sections.

Chemical synthesis

Chemical synthesis often is a good option for preparing short to medium-sized peptides (2-10 amino acids). It is also an essential tool for exploring the structure-function relationship of peptides since it enables introduction of numerous changes in the structure that will lead to modified properties. For example, non-proteinogenic amino acids can be introduced in the sequence of a polypeptide for labeling purposes [59;93] or for stabilisation [114]. Unfortunately, the synthesis of small defined peptides can be quite troublesome due to the requirement of activation and protection of amino acid building blocks. Furthermore, racemization of the building blocks during synthetic steps is frequently observed. It has been stated that "as the biosynthesis even of complicated proteins in vivo occurs within seconds or minutes, the relatively tedious classical chemical synthesis of peptides and proteins in the laboratory appears to be a somewhat hopeless enterprise, especially with regard to speed" [114]. An example of this laborious process is the first chemical synthesis of insulin, which during the early 1960s took over two years to complete [57]. Chemical peptide synthesis can be performed in three variantions: solid-phase synthesis, solution synthesis, and a hybrid approach.

Solution-phase chemical synthesis is the method of choice for the large-scale production of relatively short peptides. Its main advantage is its high flexibility: the

intermediate products can be isolated and purified after each step, deprotected and elongated to obtain larger peptides [45]. However, this technology suffers from a few inherent drawbacks: 1) occurrence of racemization during peptide activation and peptide bond formation, which reduces product purity; 2) the requirement for global protection of amino groups, carboxyl groups and side-chain functionalities, which is expensive and leads to significant yield loss during all steps; 3) the poor solubility of fully protected peptide segments, even in organic solvents, which causes slow and incomplete coupling reactions and gives high levels of by-products. Moreover, the process is less than ideal from an environmental point of view due to the use of toxic organic solvents and reagents [45]. This makes it necessary to improve these processes. For example, it is desirable to eliminate diethyl ether as a peptide precipitation agent because of the high risk of explosion in large scale operations, and dichloromethane must be substituted since it is a volatile agent that when emitted damages the ozone layer. Corrosive agents like trifluoroacetic acid (TFA), hydrofluoric acid (HF), or hydrazoic acid (HN_3) used in classical azide coupling reactions must be avoided as well [114]. One example of this is the NutraSweet process for the synthesis of aspartame [129]

Solid-phase synthesis of peptides was first described by Merrifield in 1963 [82]. It consists of attaching the carboxylic acid group of the C-terminal amino acid to a solid support and subsequently coupling to the free amino group the activated carboxylate group of an incoming N-protected amino acid, thus forming a peptide bond. Next, the N-protecting group is removed and steps are repeated until the peptide of the desired sequence and length has been synthesized [93]. Initially, this method met considerable skepticism due to the high costs of the process, the lack of adequate in-process controls, and the low purity with which the final products were generally obtained [67]. With the advent of new coupling methodologies, better support resins, and more sensitive in-process monitoring techniques, as well as the use of reverse-phase HPLC for purification, most of the original concerns about the suitability of solid-phase peptide synthesis for scale-up are no longer valid. Moreover, the reactions can be automated and the problem of solubilization of the peptide no longer exists since the growing peptide is linked to an insoluble support. Due to these characteristics, solid-phase synthesis is without doubt the technique of choice for the preparation of small amounts of oligopeptides (mg to g scale, length 2-100 amino acids) [16]. Even though solid-phase synthesis can be troublesome in practice for peptides of more than 10 residues, improved chemistry and better protocols allow the synthesis of quite long peptides. For

example, the 99 residue HIV-1 protease was made this way [59;84]. Moreover, solid-phase synthesis has become a powerful tool for preparing peptide libraries via combinatorial chemistry, as well as for making peptide arrays that can be used for the screening of peptides as receptor ligands, enzyme substrates, and enzyme inhibitors [25;132].

The hybrid approach is the production of short or medium-sized peptides by the use of solid-phase synthesis of protected fragments, which are subsequently assembled by solution-phase or solid-phase methods. The approach is advantageous for synthesizing medium-sized to large peptides or proteins (more than 100 amino acids) since it combines rapid solid-phase synthesis of pure protected fragments (up to 5-10 amino acids) with coupling reactions in solution. The final stage of the process can be fully monitored and all intermediates can be purified and characterized [67 95]. One successful example of the hybrid approach is the synthesis of an insulinotropic peptide, exenatide (39 amino acids, 4.2 kDa) which is a possible therapeutic agent for the treatment of type 2 diabetes. Because of the need of kilogram to metric ton amounts of product, the synthesis strategy was changed from solid-phase synthesis to a hybrid approach, based on the assembly of three fragments [22]

However, the disadvantage is that some racemization occurs of the residue on the C-terminal side of the peptide bond that is formed during the fragment coupling reactions. In practice, chemical fragment coupling can only be performed without racemization at the C-terminal side of a Gly or Pro residue.

Fermentative production

Ribosomal synthesis. This technology is the most suitable one for large peptides and proteins (up to several hundreds of amino acids). The starting point is a DNA fragment encoding a protein sequence, which is obtained by gene isolation, enzymatic synthesis from mRNA using reverse transcriptase, or by chemical synthesis. Making use of modern methods of cloning and gene expression, this DNA fragment is efficiently transformed into the recombinant protein of interest inside the recombinant organism. Subsequently the protein is isolated from the broth. One of the first examples of the use of this technology was the synthesis of insulin peptide chains [37]. The two chains (chain A, 21 amino acids, and chain B, 30 amino acids) were expressed in two separate bacterial strains and joined *in vitro* by using S-sulfonated derivatives to create the two disulfide bonds that hold them together. Many other examples have been

reported of bacterial strains expressing useful proteins like human growth hormone [38], interferon [127] or vaccines [139]. Despite the successful expression of several peptides and proteins, a major disadvantage of ribosomal synthesis may be the instability of the recombinant proteins, which are easily degraded by the intrinsic host proteases, especially in case of smaller peptides. Overcoming this stability issue was possible in some cases by constructing hybrid polypeptides (containing the desired protein fused with a host peptide) or by expressing a multidomain polypeptide. Proinsulin [115], neuropeptide substance P [58] and α-human atrial natriuretic peptide [70] are examples of peptides expressed as concatamers that could be subsequently converted into the desired peptides by cleavage. Other methods make use of bacterial hosts lacking certain proteases, such as the well known *E. coli* BL21 strains which are devoid of OmpT, an outer membrane protease.

Another drawback of ribosomal synthesis is the frequently observed low yield of the target peptide, especially in case of short peptides (<50 amino acids). To overcome expression problems, the use of genetically fused well-expressed tags was tested in several cases. Lee and coworkers expressed a positively charged antimicrobial target peptide in fusion with acidic residues to form a non-charged peptide that was better expressed [68], and several other tags such as maltose binding protein (MBP), glutathione-S-transferase (GST), solubility-enhancing tag (SET) were tested to enhance bacterial expression and/or facilitate purification of specific target peptides [31;94]. Good results were obtained in several cases but it is important to notice that none of these tags works equally well with each partner peptide [100].

Another strategy to increase expression of peptides is to vary the used host strain. Although *E. coli* remains the standard host, it suffers from some disadvantages, such as the inability to perform post-translational modification or the endogenous production of endotoxins that have to be separated from the desired protein before its pharmaceutical use [114]. Eukaryotic systems including mammalian, yeast and insect cells which allow specific post-translational glycosylation are becoming more accessible and less expensive and can replace bacterial hosts [39].

Despite the good results obtained in the last years, ribosomal synthesis is still the preferred method only for the synthesis of large peptides (50 to over 100 amino acids) since productivity, extraction and recovery of short peptides remains a challenge. Furthermore, a drawback of this technology is the inability (with some exceptions) of incorporation of unnatural amino acids [64]

Non-ribosomal fermentative peptide synthesis. A major class of small peptides such as antifungal and antibiotic molecules is synthesized in nature via non-ribosomal peptide synthetases (NRPSs) [133;140]. Herein, peptide assembly takes place in large multimeric enzyme complexes independently from mRNA and ribosomes. The peptide synthetase acts both as a template, conferring the required amino acid specificity and as the biosynthetic machinery for peptide synthesis. Such an NRPS complex is a modular structure composed of catalytically independent sets of domains, which are responsible for different activities, i.e. amino acid recognition, activation, coupling, modification and product release [116]. NRPSs are mainly present in bacteria, fungi and yeasts, but these multi-enzyme complexes were also described in higher eukaryotes [105]. In recent years, the attention for NRPSs was fueled by the potential to obtain peptides with novel biological activities by domain engineering. The first example of a peptide (D-Phe-Pro-DKP)(DKP is diketopiperazine) produced *in E. coli* by heterologous expression of an NRPS was reported by Grünewald and coworkers [41], but more recently several other examples were disclosed [42;125]. Watanabe and coworkers demonstrated the successful heterologous expression of the echinomicin NRPS from *Streptomyces* in *E. coli* and the synthesis of des-*N*-tetramethyl triostin A by introduction of mutations in the synthetic modules of the echinomycin NRPS [134].

A major advantage of fermentative peptide synthesis via NRPSs is the possibility to produce polypeptides that are different in size and amino acid sequence and have various modifications including *N*-methylation, cyclization, and incorporation of D-amino acids. Moreover, the use of NRPS does not require amino acid protection and the reactions can be performed in water without the use of expensive reactants or toxic chemicals.

Despite these advantages the major challenge is the generation of synthases that are specific for a desired peptide. A possible approach is the use of rational engineering to alter the action of the single domains, although progress is still limited due to the high complexity of the modular arrangement of the functional units that catalyze the successive reactions [20]. Furthermore, the low productivity of the engineered synthetases and the high costs restrict the application of this technique on large scale [48]

Chemo-enzymatic peptide synthesis

General considerations. For many years chemo-enzymatic peptide synthesis has been accompanied by alternating periods of hope and disappointment, interest and deception. Chemo-enzymatic technology has currently proven to be suitable for short sequences of less than 10 residues. In this strategy the peptide is elongated stepwise by enzymatic synthesis of peptide bonds or obtained via preparation of two or more fragments that are subsequently assembled to form the desired structure.

Compared to the classical chemical methods, enzymatic peptide synthesis has several advantages: 1) Reactions take place under very mild conditions and accompanying chemical and operational hazards are reduced; 2) The high regio-specificity of enzymes allows the use of minimally protected substrates which are often inexpensive and readily available. This can simplify the synthesis considerably by omitting intermediate protection and deprotection steps; 3) Reactions are stereospecific, allowing the use of racemic starting materials and the recycling of reactants. This is achieved by combining the synthesis with resolution, followed by racemization of the unreacted isomer and the start of a new cycle [72;117].

In nature, peptide bond formation is catalyzed by the ribozyme peptidyl transferase, which posesses full catalytic activity independent of the side chain of the amino acids being coupled. However, this highly specific activity requires various ribosomal factors including rRNA [17]. Consequently, the ribosomal peptidyltransferase activity is not suitable for practical in vitro use as a biocatalyst for peptide bond formation.

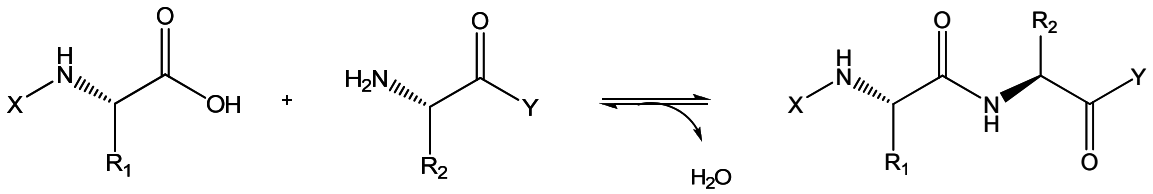
The most relevant enzymes for peptide synthesis are microbial and plant derived proteases, which are available at low costs due to their industrial production for other purposes. Proteases constitute one of the most important groups of industrial enzymes, accounting for at least a quarter of the total global enzyme production sales, and they are used in detergents, brewing, meat softening, photography, leather processing and in the pharmaceutical industry [64]. Proteases can not only catalyze peptide bond cleavage but also formation of peptide bonds. Moreover, they catalyze other types of reactions, such as esterification and transesterification, which can be used for the resolution of racemic alcohols and carboxylic acids [85], the kinetic resolution of racemic mixtures [19], and the synthesis of glycoconjugates [126]. Many proteases are good process biocatalysts since they are robust and stable enough to withstand the conditions of an industrial

process. Their pH optima are in the range of 6 to 8 and they are often active in the presence of organic solvents, supercritical fluids, ionic liquids and salts [15]. In addition, they do not require stoichiometric cofactors and may be highly stereo- and regio-selective [15].

Proteases have been widely used and currently still are the most important enzymes for peptide synthesis [49;51;64 97]. The first peptides that were synthesized with proteases were Bz-Leu-Leu-NHPh and Bz-Leu-Gly-NHPh (-NHPh is -aniline), which were made in 1938 using papain and chymotrypsin [9;10]. Another important example of chemo-enzymatic synthesis of peptides is the production of the non-caloric sweetener aspartame, which involves the use of the metalloprotease thermolysin [89]. Also the chemo-enzymatic production of human insulin starting from porcine insulin is a good example [88].

Kinetically vs. thermodynamically controlled synthesis. The equilibrium of a protease-catalyzed peptide bond formation (and cleavage) reaction is usually situated on the side of the thermodynamically more stable cleavage products. Different manipulations are possible to obtain synthesis products, in spite of the unfavourable equilibrium. These methods for peptide synthesis can be divided in two approaches: thermodynamic and kinetic control of synthesis reactions.

In the thermodynamically controlled synthesis (Scheme 2.1) the final yield of the reaction is solely determined (and limited) by the position of the thermodynamic equilibrium under the chosen reaction conditions [56]. In this process, peptide bond synthesis is the reverse of hydrolysis.



Scheme 1: Thermodynamically controlled chemo-enzymatic peptide synthesis.

The enzyme accelerates the reaction rate without effecting the equilibrium concentrations, and the formation and hydrolysis of the peptide bond proceed by the same mechanism [45]. These couplings are often done with enzymes such as thermolysin

which do not form a covalent acyl-enzyme intermediate. Various strategies have been proposed to shift the equilibrium toward synthesis:

- The use of an excess of one of the reactants;
- The removal of the product by precipitation, complexation, or extraction;
- The addition of organic solvent to the reaction mixture which can cause the pK_as of the amino and carboxyl group to approach one another;
- The pH of the reaction medium (manipulation of equilibrium of ionization);

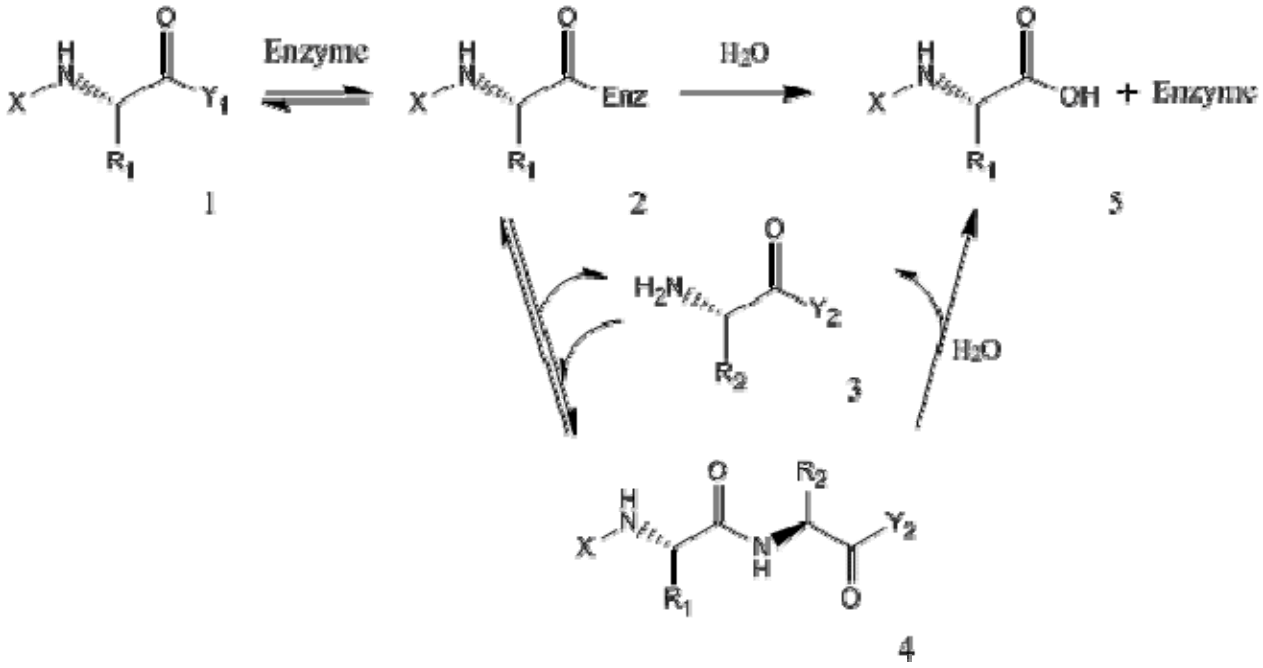
Accordingly, for each application of thermodynamically controlled peptide synthesis a study of process alternatives is needed. The choice of process conditions will depend on the reaction characteristics, on the biocatalyst, and on the properties of substrate and products. In many cases the economic embodiment of a thermodynamically controlled process is not possible.

In the kinetically controlled synthesis of peptides, the acyl-donating amino acid is activated, for example as an ester (Scheme 2.2). The protease then acts as a transferase and catalyzes the transfer of the acyl group of the activated acyl donor to the nucleophile, which in case of peptide synthesis is the amino group of the amino acid or peptide to be coupled. The reaction proceeds with transient formation of an acyl-enzyme intermediate. Besides the standard proteases (cysteine and serine proteases), a limited number of other hydrolases which are capable of forming acyl-enzyme intermediates (e.g. lipases or penicillin-G acylase) may be used for this reaction. As a side reaction, hydrolysis of the acyl-enzyme intermediate is competing with the synthetic acyl transfer reaction, leading to undersirable hydrolysis of the activated acyl donor. The synthesized peptide can also be subject to hydrolysis by the protease. The ratio of acyl transfer to nucleophile (synthesis) compared to substrate hydrolysis is an enzyme property that influences the yield [64;117]. It is therefore desirable to remove the product in order to avoid unwanted secondary hydrolysis. The product level passes through a temporary maximum, the height of which is determined by the kinetics of the biocatalyst after which the reaction proceeds (slowly) to the thermodynamic equilibrium.

The main factors that influence a kinetically controlled synthesis reaction are:

- pH. The nucleophilic group (H_2N-R) that attacks the acyl-enzyme complex must be uncharged but at the same time the ester should not be chemically hydrolyzed. The pH optimum is generally the pK_a of the amino function of the nucleophile;
- Ionic strength;

- Presence of organic (co)solvents. These help to dissolve the ester and to diminish the hydrolysis of ester and peptide;
- Enzyme specificity. The reaction rate is determined by the specificity of the enzyme for the acyl donor, but the binding of the nucleophile to the S' subsite of the protease is essential for high yields. Thus, efficiency can differ from one protease to another. The sensitivity of the acyl-enzyme intermediate towards hydrolysis and cleavage by the nucleophile is also dependent on the enzyme environment.

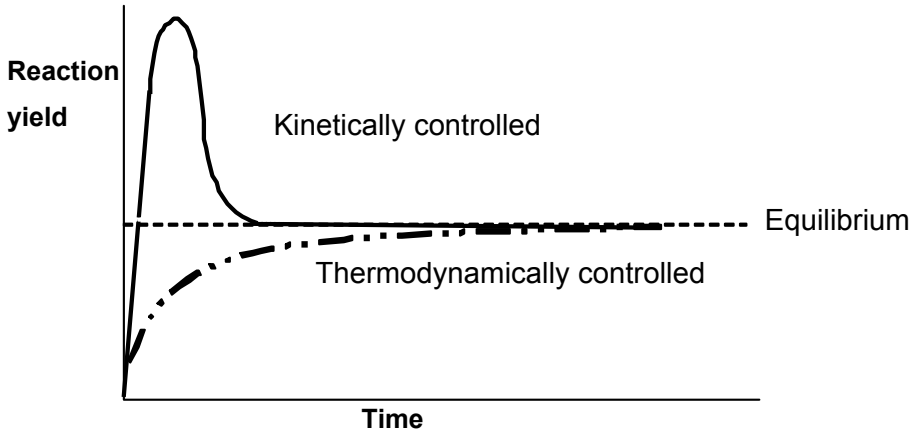


Scheme 2: Peptide synthesis by kinetic control. 1, Acyl donor; 2, acyl-enzyme complex; 3, nucleophile, 4, reaction product; 5, hydrolyzed acyl donor,

In order to achieve reasonable yields, the process has to be stopped a long time before the equilibrium is reached. This can be understood from a comparison between the time course of kinetically controlled and thermodynamically controlled reactions (Figure 2.1).

In the initial phase of the conversion, the kinetically controlled synthesis is usually rapid since the concentration of acyl donor and nucleophile are high. Once the maximum yield is reached, the effect of product hydrolysis becomes visible. Thus, the yield will decrease to finally reach the equilibrium. Additionally, part of the ester will be hydrolysed to give the free acid and nucleophile (resp. 3 and 5 in Scheme 2). The concentration of these two species and the product slowly proceed to equilibrium, so that the final yield will be determined again by the equilibrium position.

Water can have a negative effect in both approaches because it competes with the nucleophile in the kinetic approach while it influences the equilibrium position in the thermodynamic approach. Both these effects may be overcome by working in reaction media different than diluted aqueous solutions, the so-called “non conventional” media, such as organic solvents or ionic liquids.



Scheme 3: Time courses of kinetically controlled and thermodynamically controlled peptide synthesis. The dotted line represents the position of the thermodynamic equilibrium.

Direction of chemoenzymatic peptide synthesis. Several strategies which differ in the order of peptide bond formation steps can be applied for enzymatic peptide synthesis. As described above, peptides can be synthesized by a convergent strategy, coupling small fragments, or stepwise, starting from the C- or N-terminal side. When stepwise peptide synthesis is performed in the C→N direction the growing peptide chain acts as the acyl acceptor (nucleophile) during each coupling step. This way of synthesis is quite laborious since each amino acid building block needs to be *N*-protected and the peptide terminal -NH_2 must be deprotected after each step. When N→C synthesis is chosen as elongation strategy (C-terminal extension), the amino acids used as building blocks are acyl acceptors (nucleophiles) and should be added in excess of the growing peptide chain in C-protected form. A drawback in this case is that selective C-deprotection and activation is required after each step, and these reactions may need harsh conditions or expensive and toxic reagents. In this case acyl groups are activated by introduction of an electron accepting moiety, favouring the nucleophilic attack of the amino component. Many activating moieties are available, including acyl azides, carbodiimides, active esters and mixed anhydrides.

Protection of functional groups. The proper introduction and removal of protecting groups is one of the most important and widely carried out synthetic transformations. In particular in the highly selective construction of complex poly-functional molecules, a given functional group has to be selectively protected or deprotected under mild conditions, usually in the presence of functionalities of similar reactivity [54].

The selection of protecting groups is obviously dependent on the strategy applied. In fact, protecting group orthogonality is an important criterion in peptide synthesis [2], which means that a subset of protecting groups can be cleaved selectively during the synthesis of the desired molecule, while leaving all the others intact. For the manipulation of protecting groups, numerous chemical methods have been developed. Moreover, in recent years the arsenal of useful (de)protection groups has substantially increased by the application of biocatalysis [60]. In addition to their stereo-discriminating properties, enzymes offer the opportunity to carry out highly chemo- and regioselective transformations, offering viable alternatives to classical chemical methods [53].

Cα-Carboxyl protection. Protection of Cα-carboxy groups is mainly employed when peptide synthesis is conducted from the N-terminus to the C-terminus. Since this is not the usual synthetic strategy, C-terminal protection has not been widely investigated. Methyl, ethyl, benzyl or tert-butyl groups are the most used C-protecting groups. Cleavage of these protecting groups proceeds via mild alkaline (methyl and ethyl) or acidic (tert-butyl) hydrolysis or by hydrogenation (benzyl). A drawback of the use of methyl and ethyl esters as C-terminal protecting groups is that they are reactive in kinetically controlled peptide bond formation, giving product mixtures, whereas the removal of benzyl or tert-butyl groups is often accompanied by the simultaneous removal of N-protecting or side-chain protecting groups. Enzyme-catalyzed deprotection is an alternative that may solve these issues [114].

Side-chain protection. In chemical peptide synthesis it is mandatory to protect side-chain functionalities of the amino acids. The ω-protecting groups are termed “semipermanent” because they are usually cleaved off only after the last coupling step. The type of protection that is chosen largely depends on the peptide synthesis method that is used and can go from minimal to maximal protection, the latter especially in case of solid phase peptide synthesis. One of the greatest advantages of chemo-enzymatic peptide synthesis is that due to the selectivity of the coupling enzymes protection of the side chains is not needed.

N-terminal protection. More than a hundred functional groups have been developed for the protection of α-amino groups of amino acids and peptides. The most important ones are based on the acyl or carbamate function, such as the benzyloxycarbonyl ("Z" group in honor of its inventor, Leonidas Zervas), tert-butoxycarbonyl (Boc) and 9-fluorenylmethoxycarbonyl (Fmoc). Their popularity is due to their easy introduction and cleavage, which are carried out under relatively mild conditions [54;114]. A drawback of their use in C→N peptide synthesis is the need to purify each intermediate and chemically deprotect the product after each coupling step, which results in expensive procedures that are unattractive for commercial application, although racemisation hardly occurs. A better solution would be the use of an enzymatically cleavable group like formyl, which in principle could be removed by a peptide deformylase. The formyl group (For) seems industrially attractive because its introduction uses only acetic anhydride and formic acid and consequently formylation is a very cost-efficient protection method (Murakami, 2000). Its removal by acid is also cheap but gives rise to significant hydrolysis of internal peptide bonds. Hence, there is a large incentive to study the enzymatic removal of formyl groups.

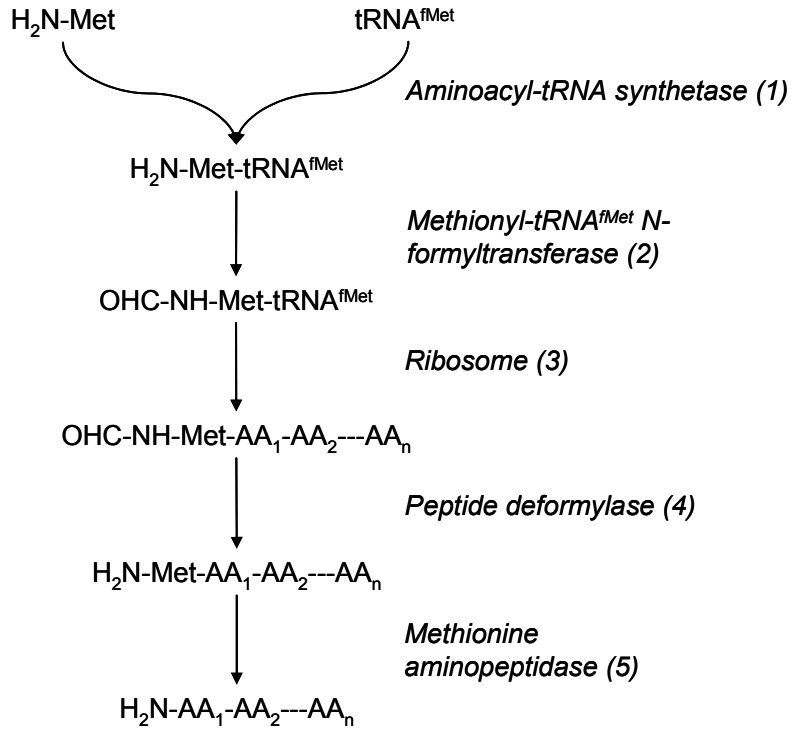
The obvious enzymes to use for the removal of formyl groups are peptide deformylases. These enzymes are generally present in prokaryotes where they are involved in the removal of formyl groups from nascent peptide chains during ribosomal protein synthesis. In prokaryotes the initiator codon on mRNA is recognized by a tRNA carrying formylmethionine, and the formyl group of the N-terminal methionine is removed after synthesis of the protein. The properties of peptide deformylases that catalyse these reactions are discussed below.

Peptide deformylases

General properties

In 1966 Capecchi observed that *N*-formyl-methionine is the unique initiator amino acid in bacterial protein synthesis. A consequence of having such common initiator is the existence of a widely distributed enzymatic cascade that results in removal of the formylmethionine, exposing other amino-terminal amino acids [18]. One year later, Adams proved for the first time the existence of an enzyme that is able to remove formate from formylmethionyl peptides in *E. coli* and *Bacillus stearothermophilus* [1]. Since then it has been discovered that the steps of tRNA binding, peptide bond formation and ribosome

translocation are virtually identical in all organisms. However, a striking difference between prokaryotes and eukaryotes is that in eubacteria as well as in mitochondria and chloroplasts, all nascent polypeptides synthesized by the ribosomal pathway initially contain an *N*-formylated methionine residue at their *N*-terminus [62], which is not the case in eukaryotes.



Scheme 4. Formylation/deformylation cycle of the prokaryotic translation initiation methionine. Step (1): aminoacyl tRNA synthetase transfers methionine to the initiating tRNA for formylmethionine. Step (2): methionyl-tRNA^{fMet} formyltransferase formylates the tRNA-bound methionine. Step (3): the formylated methionyl-tRNA^{fMet} is moved to the ribosome by the initiation factor and protein synthesis begins. Step (4): the formyl group on the N-terminal methonine of the peptide chain is removed by the activity of PDF. Step (5): the N-terminal methionine of the nascent peptide is removed by the activity of methionine aminopeptidase [69].

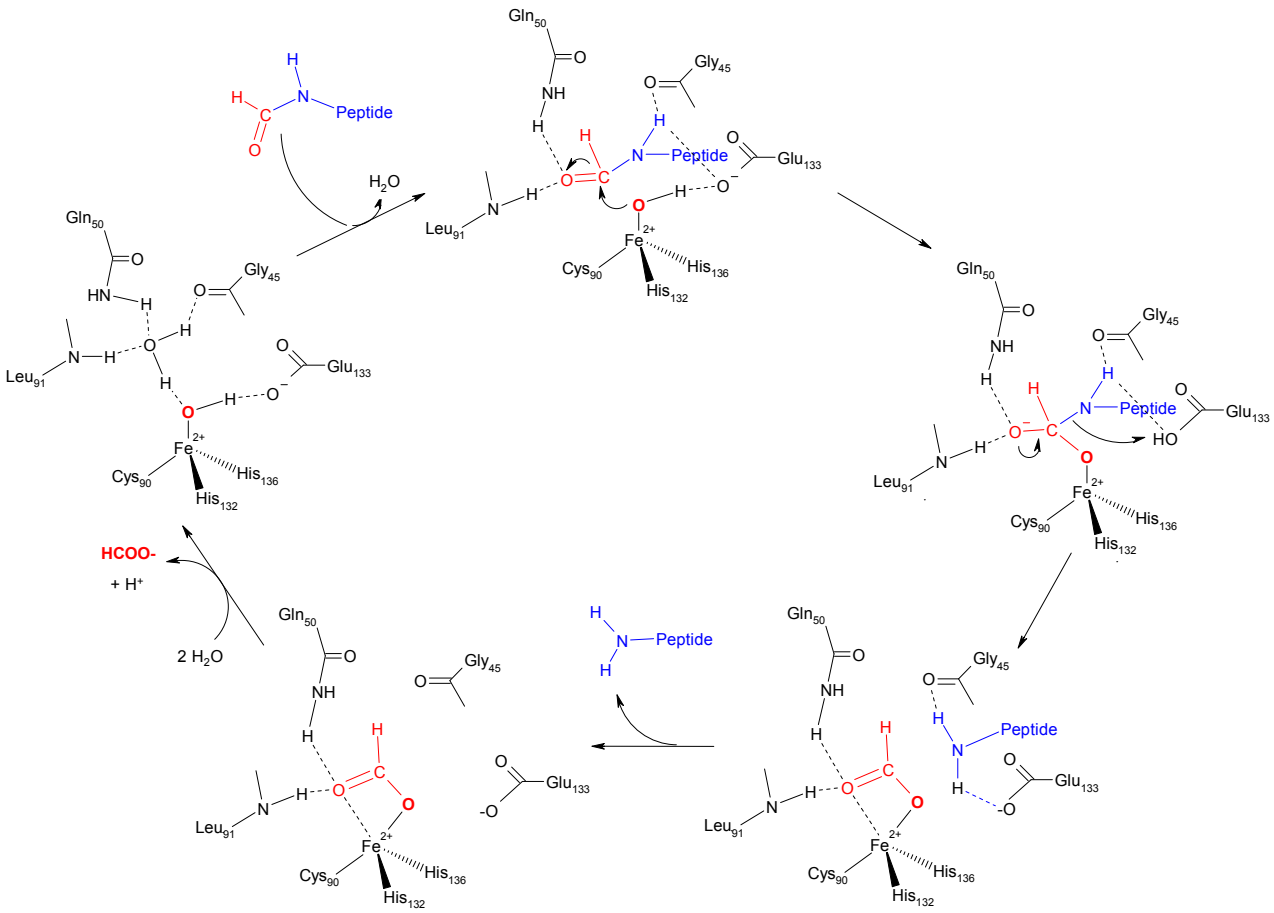
In eubacteria the transfer of the formyl group to the methionine moiety loaded to the tRNA for formylmethionine is catalyzed by methionine-tRNA^{fMet} formyltransferase (MTF, coded by the *fmt* gene), which uses 10-formyltetrahydrofolate as the formyl donor [109]. After elongation of the polypeptide chain, the formyl group is removed by the action of peptide deformylase (PDF) [1;5]. This deformylation has to take place before the N-terminal methioninecan be removed by methionine aminopeptidase (MAP) [120], which in

many cases is the second step of the maturation of the polypeptide ([14;83]. It has been reported that only 40% of the polypeptide chains in *E. coli* retain the N-terminal methionine. Furthermore, the catalytic efficiency of MAP was found to be dependent on the size of the side-chain of the penultimate amino acid [47]. Removal of this methionine can be required for the biological activity and/or stability of individual proteins [69].

Peptide deformylation is essential. Mazel *et al.* demonstrated that the inactivation of PDF, under conditions where formylation of Met-tRNA^{fMet} occurs, is lethal to *E. coli*. This supports the hypothesis that at least one essential protein requires processing by PDF and/or methionine aminopeptidase for folding or functionality [43;76]. The growth defect of peptide deformylase negative mutants could be a direct result of the formyl-blocked *N*-terminus of proteins, or it may be due to lack of removal of the methionine residue [119]. More recently, however, it was shown that *N*-terminal formylation is a dispensable feature. In fact, *def fmt* double null mutant strains, which lack both formylation and deformylation, are still able to grow, although poorly, thus adopting a process of translation initiation more similar to that of eukaryotes and Archeae. Despite the recent discoveries of active PDFs in eukaryotes, peptide deformylase is still considered an interesting antibacterial target. Thanks to the increasing knowledge on 3D structures and catalytic mechanisms, it may be possible to develop antibiotics specifically active against bacterial PDF but with little or no activity against mammalian ones [33].

PDF structure and catalytic mechanism

Structure of PDFs. Peptide deformylases (PDF; EC 3.5.1.88) belong to the superfamily of zinc metalloproteases. They contain Zn^{2+} or Fe^{2+}, the latter representing the biologically most active species. The Zn^{2+} form is more stable since it is not sensitive to oxidation. Nevertheless it was discovered that in its active form *E. coli* PDF (EcPDF) contains iron in the ferrous state (Fe^{2+}) [8]. More than 50 peptide deformylase crystal structures are available at the moment. Amino acid sequence alignments of PDFs originating from different organisms revealed a low level of overall sequence identity, but a study of the 3D structures showed a remarkably conserved topology [44;118].



Scheme 5. Reaction cycle of the Fe^{2+}-PDF. The reaction consists of 2 main steps. The first step is the nucleophilic attack of a metal-coordinated hydroxide on the carbonyl carbon of the formyl group. After formation of a tetrahedral transition-state like intermediate, the C-N bond between the formyl group and the peptide nitrogen is cleaved. The oxyanion is stabilized by side-chain groups of Gln-50 and Leu-91. The cycle is completed by peptide and formate release and reentrance of water. Adopted form Becker et al.[8]

The crystal structure of the EcPDF revealed that it is a monomeric α/β type enzyme of 168 amino acid residues composed of a single domain and containing one tightly bound metal ion [76;77;79]. The *E. coli* enzyme contains 5 α-helices and 7 β-strands organized in 3 β-sheet regions, and a potentially critical 3-10 helix. The overall fold resembles a hand composed of a five-stranded anti-parallel β-sheet and a two-stranded anti-parallel β-sheet that wraps around the central α-helix (helix II). While two of the major α-helixes (helix II and V) appear to play structural roles, residues in the third major helix (helix III) are also involved in catalysis.

The zinc ion is tetrahedrally coordinated by the Sγ-atom of Cys-90 (*E. coli* numbering), the Nε2-atom of His-132 and His-136 and a water molecule [79;80]. The two histidines are accommodated by the conserved HExxH motif in helix III, This is one of

the three conserved motifs that are distant from each other in the primary sequence but in close proximity in the 3D structure, and that are found in many zinc metallopeptidases [78]. These three motifs are residues 43-52 (*E. coli* PDF numbering) with motif 1 (sequence **G**I**G**LA**A**T***Q***V), residues 88-92 with motif 2 (***E*****G*****C*****LS)**, and the aformentioned motif 3 at residues 130-139 (**Q<u>H</u>*E***M**D<u>H</u>L**) [8;21;24;75]. The third motif contains besides the two histidines (His-132 and His-136, in bold and underlined) involved in metal ligation a glutamate residue (italics) required for catalytic activity. Other essential residues for catalysis are the glutamine of motif 1 and the glutamate of motif 2. In the latter, Glu-88 forms hydrogen bonds with His-132, Arg-102, and indirectly with Asp-135 which reinforce the enzyme structure in the vicinity of the metal ion. Moreover Glu-88 contributes to the metal ion binding via a steric effect, since its side-chain forms a lid on top of the Fe^{2+}.

The cysteine that coordinates the metal is present in motif 2 (italics). Moreover, a network of hydrogen bonding and hydrophobic interactions contributes to create a very well-defined structure around the active site.

Reaction mechanism. Based on the X-ray structure and by analyzing the structural resemblance with other metalloproteases like thermolysin, a first model of the catalytic cycle of EcPDF was proposed by Chan and co-workers in 1997 [21]. As described above, the ligands of the metal are the side chains of Cys-90, His-132, and His-136 (residing in the conserved motif HEXXH) and a bound water molecule [21].

In this model the reaction starts by substrate binding and coordination of the formyl oxygen to the zinc, resulting in formation of a five-coordinated metal. Subsequently, the water molecule shifts position to become hydrogen-bonded to both Gln-50 and Glu-133. A proton of this water is abstracted by (or hydrogen bonded to) the amide NH of the substrate, which becomes a leaving group while the deprotonated hydroxyl attacks as a nucleophile [7;24;36;79]. This first mechanism was completed by Becker *et al.* by proposing that substrate binding is accompanied by release of a water molecule that is hydrogen bonded to the water which acts as a ligand of the metal, and by attributing a role to Leu-91 (main chain NH) and Gln-50 (side chain NH) in formation of an oxyanion hole that binds the carbonyl oxygen. In agreement with this model, various substitutions at position Gln-50 and Glu-133 led to a huge decrease of the catalytic efficiency[99 102]. The reaction proceeds with the release of the peptide and the hydrolysis of the activated enzyme-formate complex releasing formate. Finally two water molecules are taken up [8;136;137].

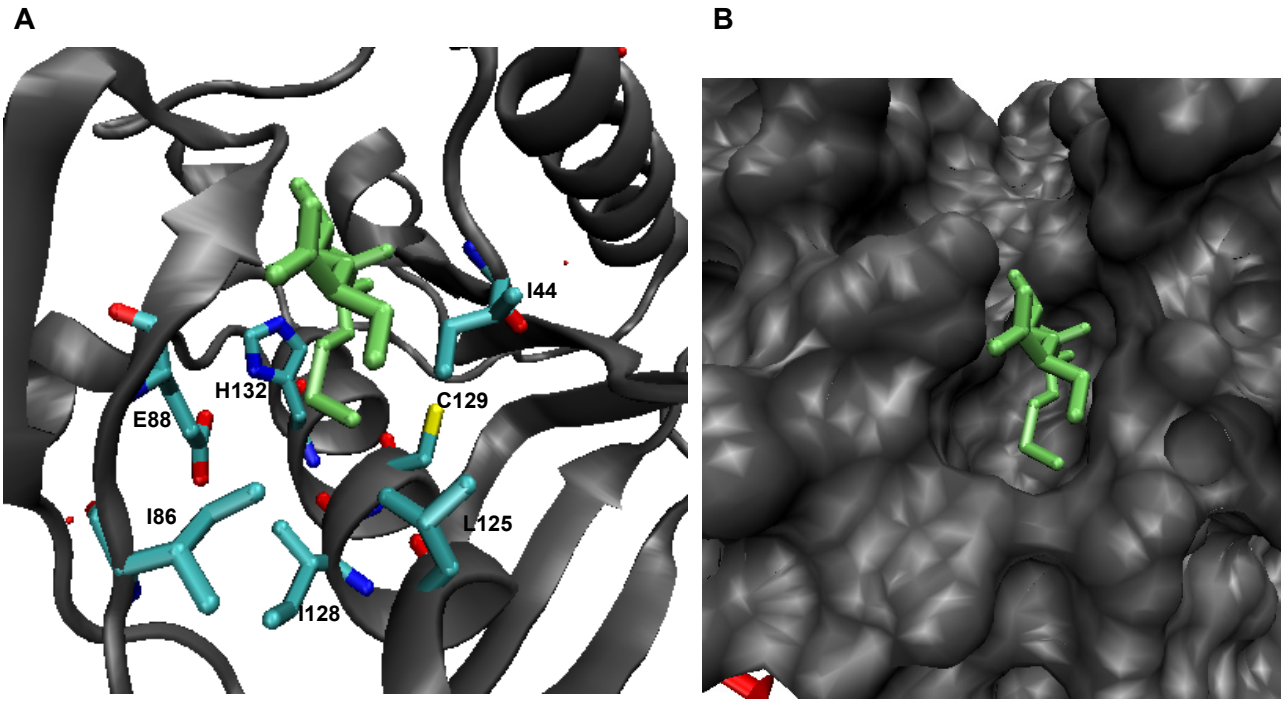


Figure 1. Close-up of the active site and pocket of EcPDF. **A**: Representation of H-Met-Ala-OH (in green) in the active site of PDF. The residues composing the hydrophobic pocket where the side chain of methionine binds are shown in cyan. **B**: Visualisation in surface mode.

Activity and specificity. The small size of the active site prevents groups larger than an N-terminal formyl from accessing the metal center [50;77]. Thus, PDF has a very high specificity for N-formylated peptides. Furthermore, the crystal structure of EcPDF in complex with H-Met-Ala-Ser-OH highlights the deep hydrophobic pocket of the S1' subsite in which the side chain of the methionine is positioned [8]. The shape of this pocket explains the general preference for hydrophobic residues at the P1' position and the dramatically lower deformylation efficiency in case of hydrophilic or charged residues [50;99].

The substrate specificity and recognition of EcPDF were comprehensively studied by Ragusa and coworkers [99] illustrating that methionine in position P1' is the best substrate for PDF compared to any other natural amino acid (Table 2). The catalytic efficiency for substrates with another N-terminal residue than methionine can be 2 to 5 orders of magnitude lower.

The S2' and S3' subsites can be described as shallow pockets, which permit different hydrophobic and bulky side chains to bind. This is in line with the natural role of PDF to accept a diverse range of peptide sequences as substrates. Furthermore, Meinnel *et al.* demonstrated that the length of the peptides also has an influence on the

activity of PDF. It was shown that the catalytic efficiency for *N*-formyl-Met-OH was about three orders of magnitude lower than for *N*-formyl-Met-Ala-OH and *N*-formyl-Met-Ala-Ser-OH [81]. For PDF from higher plants, Serero *et al.* reported a catalytic efficiency (k_{cat}/K_m) that is comparable with that of EcPDF, and the Michaelis-Menten constant (K_m) values were one order of magnitude lower. Indeed, deformylation was proven to be an essential function in chloroplasts [113].

Engineering studies. Based on analysis of the 3D structures of EcPDF [7], and on sequence alignments Ragusa and coworkers [99] proved a role for several residues in substrate recognition. To validate this, a site-directed mutagenesis approach was followed. From the stability and activity data obtained, the S1' hydrophobic pocket, accommodating the side chain of methionine (P1'), was proposed to be composed of 3 parts:

- Ceiling: framed by the side chains of residues Glu-88 and His-132
- Floor: lined by Ile-44 and Cys-129
- Bottom: lined by the side chains of Ile-86, Leu-125 and Ile-128.

The side chains of these amino acids give the cleft a hydrophobic character. As described above, among these residues only two (Glu-88 in sequence motif 2 and His-132 in motif 3 (metal binding)) are highly conserved and play an important role in the hydrogen bond network and in 3D folding. To evaluate the role played by the side chains of the other amino acids composing the pocket, several mutants were tested for their catalytic efficiency. The residues present at the bottom of the cleft, specifically Leu-125 and Ile-128, strongly influence the affinity of the enzyme for the substrate, probably due to the interaction with the methyl group of the methionine side chain. Other residues that shape the S1' pocket are not conserved despite the very similar selectivity of PDFs [63;80]. Therefore, the high specificity must be due to the overall size and hydrophobicity of the crevice. An experimental study aimed at broadening the substrate specificity of EcPDF is presented in Chapter 5 of this thesis.

Table 2: Effect of the nature of the N-terminal amino acid residue on the hydrolysis of a combinatorial peptide library by EcPDF [99].

Formylated N-terminal amino acid	k_{cat}/K_m ($M^{-1}s^{-1}x10^3$)	Relative k_{cat}/K_m
Methionine	131±11	100
Phenylalanine	37±4	28
Leucine	3.0±0.3	2.3
Tyrosine	2.4±0.3	1.8
Isoleucine	1.8±0.08	1.4
Glutamine	1.8 ±0.02	1.4
Alanine	0.84±0.08	0.6
Valine	0.19±0.02	0.14
Histidine	0.07±0.01	0.05
Lysine	0.05±0.01	0.04
Serine	0.04±0.01	0.03
Threonine	0.034±0.004	0.026
Glutamic acid	0.018±0.002	0.014
Glycine	0.017±0.002	0.013
Arginine	0.015±0.002	0.011
Asparagine	0.011±0.001	0.008
Aspartic acid	0.0009±0.00001	0.0007
Proline	«0.00003	«0.00002
Tryptophan	NS	NS

PDF stability and inactivation

As mentioned above, the first attempts to purify and characterize PDF from *E. coli* were hampered by its high instability leading to rapid inactivation of the enzyme. This prevented purification and thorough biochemical characterization of the enzyme for over 30 years. Groche *et al.* were the first to propose that this instability could be due to the involvement of a Fe^{2+} ion in the catalytic cycle of PDF which is prone to oxidation to Fe^{3+} by atmospheric O_2 and other reactive oxygen species such as H_2O_2 [40;101;103]. In 1998, another study of Meinnel and co-workers showed that PDF activity could be maintained by addition of Ni^{2+} or Mn^{2+} [98].

The inactivation of PDF was proposed to proceed in two steps, both requiring an oxidative agent, such as atmospheric oxygen. First, a ferric intermediate is generated by a one-electron oxidation by O_2. The generated superoxide can diffuse out of the active site or proceed in a second reaction, the oxidation of Cys-90, one of the metal ligands,

into cysteic acid. Moreover, the superoxide can disproportionate into H_2O_2 and O_2 and oxidize another enzyme molecule. It has been proven that only about 50% of all PDFs undergo cysteine oxidation, and that this covalent modification does not allow the recovery of activity even after incubation in the presence of Fe^{2+} or Co^{2+} [101].

In agreement with an oxidative inactivation mechanism in which peroxide plays a role, enzyme inactivation can be prevented by the combined use of catalase, for the removal of H_2O_2, and tris-(carboxyethyl)phosphine (TCEP) to arrest the first O_2-dependent step [101]. Similarly, Pei and co-workers proved that PDF activity can be maintained by exclusion of H_2O_2 and O_2 using glucose + glucose oxidase + catalase [101]. Other scavengers (like thiols or ascorbate) either have no effect or accelerate enzyme inactivation. This is due to their ability to convert O_2 in other reactive species which can oxidize the ferrous ion.

Another method that can be applied to avoid inactivation is substitution of the Fe^{2+} ion with a transition metal ion that is insensitive to oxidation, such as Ni^{2+} [40;98], Co^{2+} [102] or Mn^{2+} [98]. These modified PDF forms are active and extremely stable. It is now accepted that the Zn^{2+}-containing PDFs that are obtained by purification from recombinant PDF-expressing bacteria, must be an artefact of the overexpression and purification process. The Zn^{2+} ions present in the buffers may displace the Fe^{2+} ions, since zinc ions bind much more tightly to PDF than iron [40]. The Zn^{2+} form of the enzyme has a lower catalytic activity than the Fe^{2+} form [8]. In 1998, structural studies with EcPDF showed that in Zn-PDF one of the water molecules involved in catalysis is absent and the other is displaced, leading to a different hydrogen bonding pattern that is not optimal for catalysis [7]. Subsequent studies by Wagner and co-workers suggested that the lower catalytic activity of the Zn^{2+} form is due to its difficulty in transformation between the tetrahedral and 5-fold coordinated form during the reaction [8;40]. More recent publications explain the lower activity by a difference in interaction between the metal (Fe^{2+} or Zn^{2+}) and formate. In the case of Fe^{2+}- containing PDFs, the metal ion is bound to formate in a bidentate way, with both formate oxygen coordinated with the metal ion. In the case of Zn^{2+}-PDFs, only one of the formate oxygens is positioned at a distance that allows coordination with the formate, creating a monodentate binding. This difference in formate binding could also be the reason for the 100-fold lower activity of Zn^{2+}-containing PDFs since in these enzymes the formyl-carbonyl formed during the first step of the reaction cycle (Scheme 5) is only activated by the hydrogen bonds to the enzyme and not by the metal ion [52;55].

Diversity of PDFs

PDFs can be divided in three major subfamilies. Originally, type I included all eukaryotic PDFs and most PDFs from Gram-negative bacteria, type II mainly encompassed PDFs from Gram-positive bacteria, and type III represented PDFs coming from Archaea and trypanosomes [33;63;69]. However, the availability of crystal structures of many PDFs has made clear that the correlation between type I and II and phylogenetic position of the bacterial host is not always valid. Moreover, the analysis of several genomes showed that the number of different PDFs per organism may vary from one (*E. coli*) to four (*S. coelicolor*) and that type I and II PDFs can be present in the same organism [44]. The main difference between type I and type II is the presence of an additional stretch of amino acids at the N- and C-terminus of the type II enzymes [63].

The high conservation of the catalytic core supports the idea of common catalytic properties although more structural data about the enzyme-substrate complex will be needed to confirm this hypothesis [26]. This does not hold true in the case of higher eukaryotes, where the role of PDF has not yet been completely clarified. Human PDF (HsPDF) has been reported to exhibit *in vitro* activity towards *N*-formyl-peptides but its catalytic efficiency is very poor compared to EcPDF [30;35;91]. Another important difference between distinct peptide deformylases is their preference between zinc and other divalent cations. Despite the wide preference for zinc in many metallohydrolases, peptide deformylases predominantly prefer iron as metal ion. In several structural studies [7] the only difference at the active site of Zn^{2+}, Fe^{2+} and Ni^{2+} reported was a smaller tetrahedral volume at the zinc metal center which has been associated with tighter binding of this metal ion, which is reflected in differences in activity, as mentioned above [26].

Recently PDFs whose zinc forms are almost as active as the Fe^{2+}-containing EcPDF have been identified. These include *Arabidopsis thaliana* PDF1A (AtPDF1A) [33;113], *Leptospira interrogans* PDF (LiPDF)[71;141], *Borrelia burgdorferi* PDF (BbPDF) and *Lactobacillus plantarum* PDF (LpPDF) [92]. These enzymes share the same conserved fold as EcPDF.

Based on the activity reported for BbPDF, containing Zn^{2+} in its catalytic core, Nguyen and co-workers propose that the activity differences between Fe-EcPDF and Zn-EcPDFs can not be due to an intrinsic property of the ion but must be related to the combination of the metal and its ligand environment, creating a slightly different metal-

ligand bond length and geometry. Consequently, the two metals may bind to the same enzyme differently so that the formyl group is more or less optimally aligned in the active site for the first nucleophilic attack. This could cause the observed differences in activity [92].

Biocatalytic potential of PDF

Because of its catalytic activity, PDF can be seen as an attractive biocatalyst in chemo-enzymatic peptide synthesis. The enzymatic deprotection of formylated peptides used in a C→N synthetic strategy could in fact be performed using peptide deformylases. Although much information is available on the use of this enzyme as a target for antibacterial, antiparasitic and chemotherapeutic agents [30;69;74], data on other applications are poor.

Concerning the use of PDF in organic synthesis, the single literature information available focuses on the application in the kinetic resolution of amines and amino acid derivatives [121]. Using the native EcPDF, the authors showed that peptide deformylase could be used for the kinetic resolution of N-formylated α-amino acids and β-amino acids. The latter were converted more slowly. Also α-amino acid amides and α-amino nitriles were accepted and converted with good yield. In all cases, the (*S*)-enantiomers of the amino acid derivatives were preferentially deformylated. Very high enantioselectivities (E>300) were found with the N-amidated derivatives of phenylglycine, 3-amino-3-phenylpropionic acid, phenylglycine amide, and *tert*-leucine amide, and the catalytic rates were considered high enough for the development of a biocatalytic process. The deformylation reaction also proceeded with the dipeptide *N*-formyl-Leu-Tle-$NHCH_3$ (Tle is L-*tert*-leucine).

The EcPDF could also be used in the reverse reaction, in agreement with an earlier observation by D. Groche (1995, PhD thesis, Univ. Heidelberg). The enzyme catalyzed stereoselective formylation of α-amino acid nitriles when a high concentration of formate (thermodynamically controlled reaction) or a suitable formyl donor (kinetically controlled reaction) was used. The transformylation reaction suggested a ping-pong mechanism in which the formyl group stays bound to the enzyme after cleavage of the C-N bond in the formyl donor. The carbonyl carbon then should be cleaved by attack of the amino group of the amino acid or peptide that acts as acceptor. In agreement with the selectivity for methionine, *N*-formyl-Met-Ala-OH worked much better as a formyl donor for transfer to H-Met-Ala-Ser-OH than for formylation of phenylalanine nitrile.

Formylation of 2-amino-3-phenylpropionitrile proceeded much more efficiently with an ammonium formate solution under thermodynamic conditions, in which concentrations up to 6 M sodium formate were used [121]. These formylation reactions can be explained by the reverse of the reaction mechanism shown in Scheme 5, and the observed substrate specificities confirmed that the selectivity is mainly in the S1' pocket of EcPDF.

A disadvantage of the application of PDF in peptide synthesis is the need to rigorously separate the enzyme from peptidases and amidases that originate form the *E. coli* strain used for its overexpression because use of a crude extract will result in extensive peptide bond hydrolysis. A solution for this issue will be presented in Chapter 2 of this thesis.

Outline of this thesis

The goal of the research reported in this thesis is to explore the potential of the use of peptide deformylase in chemo-enzymatic peptide synthesis. PDF is suitable for the selective *N*-terminal deformylation of certain *N*-formyl-peptides without concurrent peptide bond hydrolysis. The challenge of applying PDF industrially prompted us to focus on the development of a new purification method to remove proteins with peptidase activity from the PDF. Other aspects such as improving the stability and modifying the substrate specificity are also investigated. In order to broaden the specificity of PDF, we decided to use saturated mutagenesis at positions that play an important role in substrate recognition. A fast assay for the rapid selection of improved peptide deformylase variants in *E. coli* colonies was developed to analyze the constructed libraries.

Chapter 2 describes the search for a cost-efficient and industrially applicable purification method. Peptide deformylase can be obtained via overexpression of the *def* gene in *E. coli* production strains. Unfortunately, its use in peptide synthesis is limited by the presence of amidases and peptidases in the cell-free extract preparation. A purification method was developed by Sonke at al. using affinity chromatography [121]. Despite the high purity obtained with this technique, the high cost of the affinity matrix is a prohibitive disadvantage for its application in an industrial setting. In order to reduce these costs, ultrafiltration and anion-exchange chromatography were explored, in combination with enzyme engineering. The first engineering approach involved the

construction of a short version of EcPDF for application in an ultrafiltration purification step. Additionally, purification via ion-exchange chromatography was facilitated by the construction of a polyglutamic acid tag, inserted on a flexible and exposed loop of peptide deformylase.

An approach to find peptide deformylases with higher stability and better catalytic properties for application in peptide synthesis is the screening of genomic databases and related information. In **Chapter 3,** eight peptide deformylases chosen from phylogenetically different classes were successfully overexpressed in *E. coli*. Partial characterization of these enzymes revealed different temperature stabilities and temperature optimum profiles but no major differences in co-solvent stability and substrate acceptance. Therefore, to obtain PDFs with tuned catalytic properties other approaches like directed evolution or rational design need to be used.

In **Chapter 4**, we discuss the advantages and pitfalls of various screening and selection methods for the discovery of the best candidates in a directed evolution library. The screening example given in this chapter is related to the growth-based selection of enzymes possessing a different substrate specificity compared with the wild-type from a library of mutants.

Chapter 5 deals with the need to obtain a peptide deformylase with broader substrate specificity for application in deprotection steps in a peptide synthesis process. The four non conserved residues known to be involved in substrate binding in the EcPDF were simultaneously randomized by site-saturated mutagenesis. The library obtained was screened for increased activity towards different substrates using a growth selection strategy. Various clones displaying new substrate specificity were recovered, sequenced and partially characterized.

In the last **Chapter 6**, the obtained results are shortly summarized and discussed with respect to their importance for the biocatalytic application of peptide deformylase.

References

1. **Adams, J. M.** 1968. On the release of the formyl group from nascent protein. J. Mol. Biol. **33**:571-574.
2. **Albericio, F.** 2000. Orthogonal protecting groups for N(alpha)-amino and C-terminal carboxyl functions in solid-phase peptide synthesis. Biopolymers **55**:123-139.
3. **Albericio, F.** 2004. Developments in peptide and amide synthesis. Curr. Opin. Chem. Biol. **8**:211-221.

4. **Anderson, C. L., C. Chaudhury, J. Kim, C. L. Bronson, M. A. Wani, and S. Mohanty**. 2006. Perspective - FcRn transports albumin: relevance to immunology and medicine. Trends Immunol. **27**:343-348.

5. **Ball, L. A. and P. Kaesberg**. 1973. Cleavage of the *N*-terminal formymlethionine residue from a bacteriophage coat protein in vitro. J. Mol. Biol. **79**:531-537.

6. **Bankowski, K., M. Manning, J. Haldar, and W. H. Sawyer**. 2002. Design of potent antagonists of the vasopressor response to arginine-vasopressin. J. Med. Chem. **21**:850-853.

7. **Becker, A., I. Schlichting, W. Kabsch, S. Schultz, and A. F. Wagner**. 1998. Structure of Peptide Deformylase and Identification of the Substrate Binding Site. J. Biol. Chem. **273**:11413-11416.

8. **Becker, A., I. Schlichting, W. Kabsch, D. Groche, S. Schultz, and A. F. V. Wagner**. 1998. Iron center, substrate recognition and mechanism of peptide deformylase. Nat. Struct. Mol. Biol. **5**:1053-1058.

9. **Bergmann, M., H. Fraenkel-Conrat,** and with the collaboration of David. 1938. The enzymatic synthesis of peptide bonds. J. Biol. Chem. **124**:1-6.

10. **Bergmann, M. and J. S. Fruton**. 1938. Some synthestic and hydrolytic experiments with chymotrypsin. J. Biol. Chem. **124**:321-329.

11. **Bickel, U., T. Yoshikawa, and W. M. Pardridge**. 2001. Delivery of peptides and proteins through the blood-brain barrier. Adv. Drug. Deliv. Rev. **46**:247-279.

12. **Biede, S. L. and E. G. Hammond**. 1979. Swiss ceese favor: I. cemical aalysis. J. Dairy Sci. **62**:227-237.

13. **Binz, H. K., P. Amstutz, and A. Pluckthun**. 2005. Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotech. **23**:1257-1268.

14. **Boesten, W., P. Raemakers-Franken, T. Sonke, G. Euverink, and P. Grijpstra**. 2003. Polypeptides having alpha-*H*-alpha amino acide amide racemase activity and nucleic acids encoding the same. United States Patent 20100292435.

15. **Bordusa, F.** 2002. Proteases in organic synthesis. Chem. Rev. **102**:4817-4868.

16. **Bruckdorfer, T., O. Marder, and F. Albericio**. 2004. From production of peptides in milligram amounts for research to multi-tons quantities for drugs of the Future. Chem. Pharm. Biotech. **5**:29-43.

17. **Burma, D.** 2001. Mechanism of peptide bond formation on ribosomes. Curr. Sci. **80**:330-331.

18. **Capecchi, M. R.** 1966. Initiation of *E. coli* proteins. Proc. Natl. Acad. Sci. USA. **55**:1517-1524.

19. **Carrera, G. and S. Riva**. 2000. Properties and synthetic applications of enzymes in organic solvents. Angew. Chem. International Edition **39**:2226-2254.

20. **Challis, G. L. and J. H. Naismith**. 2004. Structural aspects of non-ribosomal peptide biosynthesis. Curr. Opin. Struct. Biol. **14**:748-756.

21. **Chan, M. K., W. Gong, P. T. R. Rajagopalan, B. Hao, C. M. Tsai, and D. Pei**. 1997. Crystal structure of the *Escherichia coli* peptide deformylase. Biochemistry **36**:13904-13909.

22. **Chen, L., Han, Y., and Roberts, C**. 2009. Insulinotropic peptide synthesis using solid and solution phase combination techniques. European Patent EP1513946.

23. **Crombez, L., M. C. Morris, S. Dufort, G. Aldrian-Herrada, Q. Nguyen, G. Mc Master, J. L. Coll, F. Heitz, and G. Divita**. 2009. Targeting cyclin B1 through peptide-based delivery of siRNA prevents tumour growth. Nucl. Acids Res. **37**:4559-4569.

24. **Dardel, F., S. Ragusa, C. Lazennec, S. Blanquet, and T. Meinnel**. 1998. Solution structure of nickel-peptide deformylase. J. Mol. Biol. **280**:501-513.

25. **Dong-Sik, S., D. Kim, W. Chung, and Y. Lee**. 2005. Combinatorial solid phase peptide synthesis and bioassays. ChemInform **38**:517-525.

26. **Dong, M. and H. Liu**. 2008. Origins of the different metal preferences of *Escherichia coli* peptide deformylase and *Bacillus thermoproteolyticus* thermolysin: a comparative quantum mechanical/molecular mechanical study. J. Phys. Chem. B **112**:10280-10290.

27. **Edwards, C. R. W., M. J. Kitau, T. Chard, and G. M. Besser**. 1973. Vasopressin analogue DDAVP in diabetes insipidus: clinical and laboratory studies. Br. Med. J. **3**:375-378.

28. **Eichler, J.** 2008. Peptides as protein binding site mimetics. Curr. Opin. Chem. Biol. **12**:707-713.

29. **Erdmann, K., B. W. Y. Cheung, and H. Schröder**. 2008. The possible roles of food-derived bioactive peptides in reducing the risk of cardiovascular disease. J. Nutr. Biochem. **19**:643-654.

30. **Escobar-Alvarez, S., Y. Goldgur, G. Yang, O. Ouerfelli, Y. Li, and D. A. Scheinberg**. 2010. Structure and activity of human mitochondrial peptide deformylase, a novel cancer target. Mol. Cell. Biol. **30**:5099.

31. **Esposito, D. and D. K. Chatterjee** . 2006. Enhancement of soluble protein expression through the use of fusion tags. Curr. Opin. Biotechnol. **17**:353-358.

32. **Field, D., P. Connor, P. Cotter, C. Hill, and R. Ross**. 2008. The generation of nisin variants with enhanced activity against specific Gram-positive pathogens. Mol. Microbiol. **69**:218-230.

33. **Fieulaine, S., C. Juillan-Binard, A. Serero, F. Dardel, C. Giglione, T. Meinnel, and J. L. Ferrer**. 2005. The crystal structure of mitochondrial (Type 1A) peptide deformylase provides clear guidelines for the design of inhibitors specific for the bacterial forms. J. Biol. Chem. **280**:42315-42324.

34. **Futaki, S., T. Suzuki, W. Ohashi, T. Yagami, S. Tanaka, K. Ueda, and Y. Sugiura**. 2001. Arginine-rich Peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. J. Biol. Chem. **276**:5836-5840.

35. **Giglione C and Meinnel T**. 2001. Peptide deformylase as an emerging target for antiparasitic agents. Expert. Opin. Ther. Targets **5**:41-57.

36. **Giglione C, P. M. M. T.** 2000. Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. Mol. Microbiol. **36**:1197-1205.

37. **Goeddel, D. V., D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs**. 1979. Expression in Escherichia coli of chemically synthesized genes for human insulin. Proc. Natl. Acad. Sci. USA **76**:106-110.

38. **Goeddel, D. V., H. L. Heyneker, T. Hozumi, R. Arentzen, K. Itakura, D. G. Yansura, M. J. Ross, G. Miozzari, R. Crea, and P. H. Seeburg**. 1979. Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. Nature **281**:544-548.

39. **Green, J.** 2009. Host cell compatibility in protein expression. Methods Mol Biol **267**:3-14.

40. **Groche, D., A. Becker, I. Schlichting, W. Kabsch, S. Schultz, and A. F. V. Wagner**. 1998. Isolation and crystallization of functionally competent *Escherichia coli* peptide deformylase forms containing either iron or nickel in the active site. Biochem. Biophys. Res. Commun. **246**:342-346.

41. **Gruenewald, S., H. D. Mootz, P. Stehmeier, and T. Stachelhaus**. 2004. In vivo production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. Appl. Environ. Microbiol. **70**:3282-3291.

42. **Grunewald, J. and M. A. Marahiel**. 2006. Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. Microbiol. Mol. Biol. Rev. **70**:121-146.

43. **Guillon, J. M., Y. Mechulam, J. M. Schmitter, S. Blanquet, and G. Fayat**. 1992. Disruption of the gene for Met-*t*RNA(fMet) formyltransferase severely impairs growth of *Escherichia coli*. J. Bacteriol. **174**:4294-4301.

44. **Guilloteau, J. P., M. Mathieu, C. Giglione, V. Blanc, A. Dupuy, M. Chevrier, P. Gil, A. Famechon, T. Meinnel, and V. Mikol**. 2002. The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: a platform for the structure-based design of antibacterial agents. J. Mol. Biol. **320**:951-962.

45. **Guzman, F., S. Barberis, and A. Illanes**. 2007. Peptide synthesis: chemical or enzymatic. Electronic J. Biotech. **10**:279-314.

46. **Hartmann, R. and H. Meisel**. 2007. Food-derived peptides with biological activity: from research to food applications. Curr. Opin. Biotechnol. **18**:163-169.

47. **Hirel, P. H., M. J. Schmitter, P. Dessen, G. Fayat, and S. Blanquet**. 1989. Extent of *N*-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA **86**:8247-8251.

48. **Hodgson D.R.W and J.M. Sanderson**. 2004. The synthesis of peptides and proteins containing non-natural amino acids. Chem. Soc. Rev. **33**:422-430.

49. **Hou, R. Z., N. Zhang, G. Li, Y. B. Huang, H. Wang, Y. P. Xiao, Y. J. Liu, Y. Yang, L. Zhao, and X. Z. Zhang**. 2005. Synthesis of tripeptide RGD amide by a combination of chemical and enzymatic methods. J. Mol. Catal. B: Enzymatic **37**:9-15.

50. **Hu, Y. J., Y. Wei, Y. Zhou, P. T. R. Rajagopalan, and D. Pei**. 1999. Determination of substrate specificity for peptide deformylase through the screening of a combinatorial peptide library. Biochemistry **38**:643-650.

51. **Huang, Y. B., Y. P. Xiao, H. Wang, R. Z. Hou, N. Zhang, X. X. Wu, L. Xu, and X. Z. Zhang**. 2005. Chemo-enzymatic synthesis of tripeptide RGD diamide in organic solvents. J. Biotech. **116**:51-59.

52. **Jain, R., B. Hao, R. Liu, and M. K. Chan**. 2005. Structures of *E. coli* peptide deformylase bound to formate: insight into the preference for Fe^{2+} over Zn^{2+} as the active site metal. J. Am. Chem. Soc. **127**:4558-4559.

53. **Kadereit D., J. Kuhlmann, and H. Waldmann**. 2000. Linking the fields - The interplay of organic synthesis, biophysical chemistry and cell biology in the chemical biology of protein lipidation. Chem. Bio.Chem. **1**:144-169.

54. **Kadereit, D. and H. Waldmann**. 2001. Enzymatic protecting group techniques. Chem. Rev. **101**:3367-3396.

55. **Karambelkar, V. V., C. Xiao, Y. Zhang, A. A. NarducciSarjeant, and D. P. Goldberg**. 2006. Geometric Preferences in iron(II) and zinc(II) model complexes of peptide deformylase. Inorg. Chem. **45**:1409-1411.

56. **Kasche, V.** 1986. Mechanism and yields in enzyme catalysed equilibrium and kinetically controlled synthesis of [beta]-lactam antibiotics, peptides and other condensation products. Enzyme and Microbial Technol. **8**:4-16.

57. **Katsoyannis, P. G.** 1966. The chemical synthesis of human and sheep insulin. The Am. J. Med. **40**:652-661.

58. **Kempe, T., S. B. H. Kent, F. Chow, S. M. Peterson, W. I. Sundquist, J. J. L'Italien, D. Harbrecht, D. Plunkett, and W. J. DeLorbe**. 1985. Multiple-copy genes: production and modification of monomeric peptides from large multimeric fusion proteins. Gene **39**:239-245.

59. **Kent, S.** 2008. Total chemical synthesis of proteins. Chem. Soc. Rev. **38**:338-351.

60. **Kocienski, P. J.** 2005. Protecting Groups, 3rd edition: Thieme

61. **Korhonen, H. and A. Pihlanto**. 2006. Bioactive peptides: Production and functionality. Int. Dairy J. **16**:945-960.

62. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Mol. Biol. Rev. **47**:1-45.

63. **Kreusch, A., G. Spraggon, C. C. Lee, H. Klock, D. McMullan, K. Ng, T. Shin, J. Vincent, I. Warner, C. Ericson, and S. A. Lesley**. 2003. Structure analysis of peptide deformylases from *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Thermotoga maritima* and *Pseudomonas aeruginosa*: Snapshots of the oxygen sensitivity of peptide deformylase. J. Mol. Biol. **330**:309-321.

64. **Kumar, D. and T. Bhalla**. 2005. Microbial proteases in peptide synthesis: approaches and applications. Appl. Microbiol. Biotechnol. **68**:726-736.

65. **Landon, L. A., J. Zou, and S. L. Deutscher**. 2004. Is phage display technology on target for developing peptide-based cancer drugs? Curr. Drug Discov. Tech. **1**:113-132.

66. **Langer, M., F. Kratz, B. Rothen-Rutishauser, H. Wunderli-Allenspach, and A. G. Beck-Sickinger**. 2001. Novel petide cnjugates for tmor-secific cemotherapys. J. Med. Chem. **44**:1341-1348.

67. **Andersson, L., L. Blomberg, M. Flegel, L. Lepsa, B. Nilsson, and M. Verlander**. 2000. Large-scale synthesis of peptides. Peptide science **55**:227-250.

68. **Lee, J. H., I. Minn, C. B. Park, and S. C. Kim**. 1998. Acidic pptide-mdiated epression of the atimicrobial pptide buforin II as tandem repeats in *Escherichia coli*. Protein Express. Purif. **12**:53-60.

69. **Leeds, J. A. and C. R. Dean**. 2006. Peptide deformylase as an antibacterial target: a critical assessment. Curr. Opin. Pharmac. **6**:445-452.

70. **Lennick, M., J. R. Haynes, and S. H. Shen**. 1987. High-level expression of [alpha]-human atrial natriuretic peptide from multiple joined genes in *Escherichia coli*. Gene **61**:103-112.

71. **Li, Y., Z. Chen, and W. Gong**. 2002. Enzymatic properties of a new peptide deformylase from pathogenic bacterium *Leptospira interrogans*. Biochem. Biophys. Res. Commun. **295**:884-889.

72. **Lombard, C., J. Saulnier, and J. M. Wallach**. 2005. Recent trends in protease-catalyzed peptide synthesis. Prot. Pep. Letters **12**:621-629.

73. **Malkoski, M., S. G. Dashper, N. M. O'Brien-Simpson, G. H. Talbo, M. Macris, K. J. Cross, and E. C. Reynolds**. 2001. Kappacin, a novel antibacterial peptide from bovine Milk. Antimicrob. Agents Chemother. **45**:2309-2315.

74. **Mannervik, B., A. Runarsdottir, and S. Kurtovic**. 2009. Multi-substrateΓÇôactivity space and quasi-species in enzyme evolution: Ohno's dilemma, promiscuity and functional orthogonality. Biochem. Society Transactions **037**:740-744.

75. **Mazel, D., E. Coic, S. Blanchard, W. Saurin, and P. Marliere**. 1997. A survey of polypeptide deformylase function throughout the eubacterial lineage. J. Mol. Biol. **266**:939-949.

76. **Mazel, D., S. Pochets, and P. Marliere**. 1994. Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. **13**:914-923.

77. **Meinnel, T. and S. Blanquet**. 1995. Enzymatic properties of *Escherichia coli* peptide deformylase. J. Bacteriol. **177**:1883-1887.

78. **Meinnel, T., S. Blanquet, and F. Dardel**. 1996. A new subclass of the zinc metalloproteases superfamily revealed by the solution structure of peptide deformylase. J. Mol. Biol. **262**:375-386.

79. **Meinnel, T., C. Lazennec, and S. Blanquet**. 1995. Mapping of the active site zinc ligands of peptide deformylase. J. Mol. Biol. **254**:175-183.

80. **Meinnel, T., C. Lazennec, S. Villoing, and S. Blanquet**. 1997. Structure-function relationships within the peptide deformylase family. Evidence for a conserved architecture of the active site involving three conserved motifs and a metal ion. J. Mol. Biol. **267**:749-761.

81. **Meinnel, T., L. Patiny, S. Ragusa, and S. Blanquet**. 1999. Design and synthesis of substrate analogue inhibitors of peptide deformylase. Biochemistry **38**:4287-4295.

82. **Merrifield, R. B.** 1963. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J. Am. Chem. Soc. **85**:2149-2154.

83. **Miller, C. G., K. L. Strauch, A. M. Kukral, J. L. Miller, P. T. Wingfield, G. J. Mazzei, R. C. Werlen, P. Graber, and N. R. Movva**. 1987. *N*-terminal methionine-specific peptidase in *Salmonella typhimurium*. Proc. Natl. Acad. Sci. USA **84**:2718-2722.

84. **Milton, R. C., S. C. F. Milton, and S. B. H. Kent**. 1992. Total chemical synthesis of a *D*-enzyme: the enantiomers of HIV-1 protease show demonstration of reciprocal chiral substrate specificity. Science **256**:1445-1448.

85. **Miyazawa, T., K. Imagawa, H. Minowa, T. Miyamoto, and T. Yamada**. 2005. Resolution of non-protein amino acids via the microbial protease-catalyzed enantioselective hydrolysis of their *N*-unprotected esters. Tetrahedron **61**:10254-10261.

86. **Moller N.P, K.E. Scholz-Ahrens, N. Roos, and J. Schrezenmeir**. 2008. Bioactive peptides and proteins from foods: indication for health effects. Eur. J. Nutr. **47**:171-82.

87. **Monnai, O. and A. Hosono**. 1992. Bovinek-casein as inhibitor of the proliferation of mouse splenocytes induced by lipopolysaccharide stimulation. Milchwissenschaften **47**:512.

88. **Morihara, K.** 1990. Enzymatic semisynthesis of human insulin: an update. J. Mol. Recognit. **3**:181-186.

89. **Murakami Y., T. Yoshida, S. Hayashi, and A. Hirata**. 2000. Continuous enzymatic production of peptide precursor in aqueous/organic biphasic medium. Biotechnol. Bioeng. **69**:57-65.

90. **Nguyen, A., A. E. Reyes, II, M. Zhang, P. McDonald, W. L. Wong, L. A. Damico, and M. S. Dennis**. 2006. The pharmacokinetics of an albumin-binding Fab (AB Fab) can be modulated as a function of affinity for albumin. Protein Eng. Design Selection **19**:291-297.

91. **Nguyen, K. T., X. Hu, C. Colton, R. Chakrabarti, M. X. Zhu, and D. Pei**. 2003. Characterization of a human peptide deformylase: implications for antibacterial drug design. Biochemistry **42**:9952-9958.

92. **Nguyen, K. T., J. C. Wu, J. A. Boylan, F. C. Gherardini, and D. Pei**. 2007. Zinc is the metal cofactor of *Borrelia burgdorferi* peptide deformylase. Arch. Biochem. Biophys. **468**:217-225.

93. **Nilsson, B. L., M. B. Soellner, and R. T. Raines**. 2005. Chemical synthesis of proteins. Annu. Rev. Biophys. Biomol. Struct. **34**:91.

94. **Nilsson, J., S. Stσhl, J. Lundeberg, M. Uhlθn, and P. σ. Nygren**. 1997. Affinity fusion strategies for detection, purification and immobilization of recombinant proteins. Prot. Express. Purif. **11**:1-16.

95. **Nishiuchi, Y., H. Nishio, T. Inui, J. Bódi, and T. Kimura**. 2000. Combined solid-phase and solution approach for the synthesis of large peptides or proteins. J. Peptide Sci. **6**:84-93.

96. **Powell, M., T. Stewart, L. Otvos, L. Urge, F. Gaeta, A. Sette, T. Arrhenius, D. Thomson, K. Soda, and S. Colon**. 1993. Peptide stability in drug development. II. Effect of single amino acid substitution and glycosylation on peptide reactivity in human serum. Pharm. Res. **10**:1268-1273.

97. **Quiroga, E., N. Priolo, D. Obregon, J. Marchese, and S. Barberis**. 2008. Peptide synthesis in aqueous-organic media catalyzed by proteases from latex of *Araujia hortorum* (Asclepiadaceae) fruits. Biochem. Eng. J. **39**:115-120.

98. **Ragusa, S., S. Blanquet, and T. Meinnel**. 1998. Control of peptide deformylase activity by metal cations. J. Mol. Biol. **280**:515-523.

99. **Ragusa, S., P. Mouchet, C. Lazennec, V. Dive, and T. Meinnel**. 1999. Substrate recognition and selectivity of peptide deformylase. similarities and differences with metzincins and thermolysin. J. Mol. Biol. **289**:1445-1457.

100. **Raingeaud, J., F. Lavergne, V. Lelievre, J. M. Muller, R. Julien, and Y. Cenatiempo**. 1996. Production, analysis and bioactivity of recombinant vasoactive intestinal peptide analogs. Biochimie **78**:14-25.

101. **Rajagopalan, P. T. and D. Pei**. 1998. Oxygen-mediated Inactivation of peptide deformylase. J. Biol. Chem. **273**:22305-22310.

102. **Rajagopalan, P. T. R., S. Grimme, and D. Pei**. 2000. Characterization of cobalt(II)-substituted peptide deformylase: function of the metal ion and the catalytic residue Glu-133. Biochemistry **39**:779-790.

103. **Rajagopalan, P. T. R., X. C. Yu, and D. Pei**. 1997. Peptide deformylase: a new type of mononuclear iron protein. J. Am. Chem. Soc. **119**:12418-12419.

104. **Rejmann, J., M. Lewis, and S. Oliver**. 1993. Enhancement of mammary gland mononuclear cell proliferation by interleukin-2 in the presence of lactoferrin. Food Agric. Immunol. **5**:51.

105. **Richardt, A., T. Kemme, S. Wagner, D. Schwarzer, M. A. Marahiel, and B. T. Hovemann**. 2003. Ebony, a novel nonribosomal peptide synthetase for {beta}-alanine conjugation with biogenic amines in Drosophila. J. Biol. Chem. **278**:41160-41166.

106. **Rival, S. G., S. Fornaroli, C. G. Boeriu, and H. J. Wichers**. 2000. Caseins and casein hydrolysates. 1. Lipoxygenase inhibitory properties. J. Agricult. Food Chem. **49**:287-294.

107. **Sato, A. K., M. Viswanathan, R. B. Kent, and C. R. Wood**. 2006. Therapeutic peptides: technological advances driving peptides into development. Curr. Opin. Biotechnol. **17**:638-642.

108. **Schafmeister, C. E., J. Po, and G. L. Verdine**. 2000. An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. J. Am. Chem. Soc. **122**:5891-5892.

109. **Schmitt, E., J. M. Guillon, T. Meinnel, Y. Mechulam, F. Dardel, and S. Blanquet**. 1996. Molecular recognition governing the initiation of translation in *Escherichia coli*. Biochimie **78**:543-554.

110. **Schrezenmeir, J. and M. de Vrese**. 2001. Probiotics, prebiotics, and synbiotics--approaching a definition. Am. J. Clin. Nutr. **73**:361S-364.

111. **Sebbage, V.** 2009. Cell-penetrating peptides and their therapeutic applications. Bioscience Horizons **2**:64-72.

112. **Seppo, L., T. Jauhiainen, T. Poussa, and R. Korpela**. 2003. A fermented milk high in bioactive peptides has a blood pressure-lowering effect in hypertensive subjects. Am. J. Clin. Nutr. **77**:326-330.

113. **Serero, A., C. Giglione, and T. Meinnel**. 2001. Distinctive features of the two classes of eukaryotic peptide deformylases. J. Mol. Biol. **314**:695-708.

114. **Sewald, N. and H.-D. Jakubke**. 2009. Peptides: chemistry and biology, 2nd Ed.

115. **Shen, S. H.** 1984. Multiple joined genes prevent product degradation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:4627-4631.

116. **Sieber, S. A. and M. A. Marahiel**. 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. **105**:715-738.

117. **Sinisterra, J. V. and A. R. Alcantara**. 1993. Synthesis of peptides catalysed by enzymes: a practical overview. J. Mol. Catal.**84**:327-364.

118. **Smith, K. J., C. M. Petit, K. Aubart, M. Smyth, E. McManus, J. Jones, A. Fosberry, C. Lewis, M. Lonetto, and S. B. Christensen**. 2003. Structural variation and inhibitor binding in polypeptide deformylase from four different bacterial species. Protein Sci **12**:349-360.

119. **Solbiati, J., A. Chapman-Smith, J. L. Miller, C. G. Miller, and J. E. Cronan**. 1999. Processing of the *N*-termini of nascent polypeptide chains requires deformylation prior to methionine removal. J. Mol. Biol. **290**:607-614.

120. **Solbiati, J., A. Chapman-Smith, J. L. Miller, C. G. Miller, and J. E. Cronan Jr**. 1999. Processing of the *N*-termini of nascent polypeptide chains requires deformylation prior to methionine removal. J. Mol. Biol. **290**:607-614.

121. **Sonke, T., B. Kaptein, A. F. V. Wagner, P. J. L. M. Quaedflieg, S. Schultz, S. Ernste, A. Schepers, J. H. M. Mommers, and Q. B. Broxterman**. 2004. Peptide deformylase as biocatalyst for the synthesis of enantiomerically pure amino acid derivatives. J Mol Catal B: Enzymatic **29**:265-277.

122. **Stead, R. B., J. Lambert, D. Wessels, J. S. Iwashita, K. K. Leuther, K. W. Woodburn, P. J. Schatz, D. M. Okamoto, R. Naso, and A. M. Duliege**. 2006. Evaluation of the safety and pharmacodynamics of Hematide, a novel erythropoietic agent, in a phase 1, double-blind, placebo-controlled, dose-escalation study in healthy volunteers. Blood **108**:1830-1834.

123. **Steffansen, B., C. U. Nielsen, and S. Frokjaer**. 2005. Delivery aspects of small peptides and substrates for peptide transporters. Eur. J. Pharmaceut. Biopharmaceut. **60**:241-245.

124. **Sugano, M., S. Goto, Y. Yamada, K. Yoshida, Y. Hashimoto, T. Matsuo, and M. Kimoto**. 1990. Cholesterol-lowering activity of various undigested fractions of soybean protein in rats. J. Nutr. **120**:977-985.

125. **Tabata, K. and S. I. Hashimoto**. 2007. Fermentative production of *L*-Alanyl-*L*-Glutamine by a metabolically engineered *Escherichia coli* strain expressing *L*-amino acid {alpha}-ligase. Appl. Environ. Microbiol. **73**:6378-6385.

126. **Taniguchi, N., A. Suzuki, Y. Ito, H. Narimatsu, and T. Kawasaki**. 2008. Experimental glycoscience glycochemistry. Springer Japan.

127. **Taniguchi, T., L. Guarente, T. M. Roberts, D. Kimelman, J. Douhan, and M. Ptashne**. 1980. Expression of the human fibroblast interferon gene in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **77**:5230-5233.

128. **Thayer, A.** 2011. Making peptides at large scale. Chem Eng. News **89**:21-25.

129. **Tsau, J., D. Seagle, and L. Steven**. 1992. Process for preparing aspartame coated organic acid. US Patent 5114726.

130. **Veronese, F. M. and J. M. Harris** . 2008. Peptide and protein PEGylation III: advances in chemistry and clinical applications. Adv. Drug Deliv. Rev. **60**:1-2.

131. **Vlieghe, P., V. Lisowski, J. Martinez, and M. Khrestchatisky**. 2009. Synthetic therapeutic peptides: science and market. Drug. Discovery Today **15**:40-56.

132. **Stadler, V. D., D. Thomas Felgenhauer, D. Mario Beyer, Dr. Simon Fernandez, Klaus Leibe, Dr. Stefan Güttler, Martin Gröning, Kai König, Dr. Gloria Torralba, P. Michael Hausmann, P. Volker Lindenstruth, D. Alexander Nesterov, Ines Block, D. Rüdiger Pipkorn, P. Annemarie Poustka, Dr. F.Ralf Bischoff, and Dr. Frank Breitling**. 2008. Combinatorial synthesis of peptide arrays with a laser printer. Angew. Chem. International Edition **47**:7132-7135.

133. **Walsh, C.** 2003. Antibiotics: Actions, origins, resistance. ASM Press: Washington DC.

134. **Watanabe, K., H. Oguri, and H. Oikawa**. 2009. Diversification of echinomycin molecular structure by way of chemoenzymatic synthesis and heterologous expression of the engineered echinomycin biosynthetic pathway. Curr. Opin. Chem. Biol. **13**:189-196.

135. **Wong, C., H. Seow, A. Husband, G. Regester, and D. Watson**. 1997. Effects of purified bovine whey factors on cellular immune functions in ruminants. Vet. Immunol. Immunopathol. **56**:85.

136. **Wu, X. H., J. M. Quan, and Y. D. Wu**. 2007. Theoretical study of the catalytic mechanism and metal-ion dependence of peptide deformylase. J. Phys. Chem. B **111**:6236-6244.

137. **Xiao, C. and Y. Zhang**. 2007. Catalytic mechanism and metal specificity of bacterial peptide deformylase: a density functional theory QM/MM Study. J. Phys. Chem. B **111**:6229-6235.

138. **Yamamoto, N.** 1997. Antihypertensive peptides derived from food proteins. Biopolymers **43**:129.

139. **Yelverton, E., S. Norton, J. F. Obijeski, and D. V. Goeddel**. 1983. Rabies virus glycoprotein analogs: biosynthesis in *Escherichia coli*. Science **219**:614-620.

140. **Zasloff, M.** 2002. Antimicrobial peptides of multicellular organisms. Nature **415**:389-395.

141. **Zhou, Z., X. Song, Y. Li, and W. Gong**. 2004. Unique structural characteristics of peptide deformylase from pathogenic bacterium *Leptospira interrogans*. J. Mol. Biol. **339**:207-215.

Chapter 2

Purification of recombinant E. coli peptide deformylase

Abstract

Peptide deformylases (PDFs) catalyze the removal of the N-terminal formyl group from the N-terminal methionine in nascent polypeptide chains in prokaryotes. This activity makes PDFs attractive candidates for the biocatalytic deprotection of formylated peptides which can be used in chemoenzymatic peptide synthesis. PDFs can also be applied for the synthesis of enantiomerically pure amino acid derivatives. For application in peptide deprotection, it is essential to use PDF preparations that are free of contamination by amidases and peptidases since their presence would cleave internal bonds in the peptide that is under synthesis. To be able to explore the use of E. coli PDF (EcPDF) in these biocatalytic applications, a simple and industrially applicable purification procedure was developed based on a single anion-exchange chromatography step. To achieve separation, the *E. coli* PDF was engineered with an anionic tag composed of 8 glutamic acid residues. The activity and stability of the engineered enzyme were similar to those of the wild-type PDF. This purification furnished a PDF preparation with a 1,500-fold decreased level of contamination by amidases and peptidases as compared to cell-free extract. It was shown that the enzyme could be used for deprotection of a formylated dipeptide that was prepared by thermolysin-mediated coupling.

Introduction

Chemical peptide synthesis, either in solution or by solid phase methods, generally proceeds by elongation at the *N*-terminal side of the nascent peptide chain. In academic practice *N*-protecting groups like tert-butoxycarbonyl (Boc), benzyloxy-carbonyl (Cbz) and especially fluorenylmethoxycarbonyl (Fmoc) are used in combination with advanced coupling reagents like carbodiimides, phosphonium salts, or uronium/guanidinium salts. Examples of coupling reagents are 1-ethyl-3-(3-dimethyl-aminopropyl) carbodiimide *N*-hydroxysuccinimide (EDC/HOSu), 1-ethyl-3-(3-dimethyl-aminopropyl) *N*-hydroxybenzotriazole (EDC/HOBt) or benzotriazolyloxy-tris-(pyrrolidino)-phosphonium hexafluorophosphate (ByBOP) [11]. Although these coupling methods work well on a laboratory scale and proceed without substantial racemisation of the amino acid building blocks, commercial application requires cheaper protecting groups and coupling reagents. In this respect, a particularly interesting amino protecting

group is the formyl moiety, which can be readily introduced at low cost using formic acid and acetic acid anhydride [26]. The removal of the formyl group after the coupling reaction can be performed either chemically or enzymatically. Acid hydrolysis in water and/or methanol is a conventional method for removing formyl groups from the amino group of amino acids and the amino terminus of peptides [35]. However, the low pH and harsh reaction conditions lead to significant peptide bond hydrolysis when this method is applied in peptide synthesis. A mild enzymatic method would thus be more attractive.

One class of enzymes that can be used for the removal of formyl-protecting groups consists of the peptide deformylases. The cellular ribosome-mediated synthesis of proteins starts with a methionine residue. In eubacteria and eukaryotic organelles (mitochondria and chloroplasts), the methionyl moiety carried by the initiator tRNA.fMet is *N*-formylated prior to its incorporation into a polypeptide [28]. Following initiation of translation, the enzyme peptide deformylase (PDF, EC 3.5.1.88) cleaves the formyl group from the nascent polypeptide chain [2]. Next, methionyl aminopeptidase may remove the *N*-terminal methionine from the deformylated polypeptide, leading to a mature protein [8]. Because of its deformylase activity, PDF could be an attractive biocatalyst for the deprotection of formylated peptides.

Although much information is available on PDFs as a target for antibacterial, antiparasitic and chemotherapeutic agents [15;22;23], data on biocatalytic applications are scarce. Concerning the use of PDF in organic synthesis, the only information available in the literature describes the application of the enzyme in the synthesis of enantiopure amines and amino acid derivatives, based on its ability to stereoselectively hydrolyse *N*-formyl α-aminonitriles, and *N*-formyl derivatives of non-functionalized amines and β-amino alcohols [32]. In this paper, the authors also describe an example of the use of PDF for the selective deprotection of *N*-formyl-dipeptides. A prerequisite of the application of PDFs in peptide synthesis is the possibility to isolate the enzyme in a form that is free of contaminating peptidases and amidases that may originate form the *E. coli* host strain used for overexpression. The presence of peptidases would result in peptide bond hydrolysis in the peptide that is under synthesis.

The aim of the research described here is to obtain a method for the rapid, upscalable, and cheap production of purified PDF that is capable of removing *N*-terminal formyl groups from peptides without concurrent peptide bond hydrolysis. A one-step purification method for PDF based on affinity chromatography with a Met-Lys modified Sepharose matrix was described before [32]. This strategy is in practice only

suitable for use on laboratory scale since the matrix is not commercially available and its preparation involves a complicated chemo-enzymatic preparation process making use of expensive materials. In this Chapter, we describe the development of two simple and industrially applicable purification methods for EcPDF.

Because EcPDF is a monomeric enzyme of 168 aa (19.24 kDa), which is rather small as compared to the average *E. coli* amidases and peptidases, the use of ultra-filtration methods was first explored. Membrane ultrafiltration (UF) is a pressure-modified, convective process that uses semi-permeable membranes to separate species in aqueous solutions by molecular size, shape and/or charge [14]. The surface of the UF membrane contains pores with diameters small enough to distinguish between the sizes and shapes of dissolved molecules. Those above a predetermined size range are rejected, whereas those below that range pass trough the membrane with the solvent flow [14].

A second option that is described makes use of ion-exchange chromatography (IEC) as a single purification step, combined with the use of negatively charged tags fused onto specific enzyme domains. Anion exchange chromatography is a widely used protein separation technique. By improving the selectivity of IEC through modification of the target protein, this method could be a cheap and robust alternative to the traditional affinity systems. One way to do so is to change the charge distribution on the surface to allow for stronger absorption to the ion exchanger [21]. To apply this principle for the purification of EcPDF, we decided to insert a poly-anionic sequence containing eight glutamic acid residues in a solvent exposed loop on the surface of EcPDF. This allowed efficient purification using ion-exchange chromatography.

Materials and methods

Materials

All standard chemicals were of the highest grade obtainable. H-Phe-NH_2, H-TyrOMe, *N*-Formyl-Met-Ala-OH, H-Met-Ala-OH, *N*-formyl-Met-Lys-OH, H-Leu-Phe-OH, H-Phe-Leu-OH, H-Asp-Phe-OH, H-Phe-Asp-OH, H-Arg-Phe-OH, H-Phe-Arg-OH, H-Gly-Phe-OH, and H-Phe-Gly-OH were purchased from Bachem (Bubendorf, Switzerland); tris-(2-carboxyethyl)-phosphine (TCEP) was obtained from Fluka (Buchs, Switzerland). If desired, amino acid derivatives were formylated in a refluxing mixture of formic acid and 1 eq. of acetic acid anhydride.

The mono Q HR 5/5 and HiLoad Q sepharose anion exchange columns were purchased from GE Healthcare Bio-Sciences Ltd (United Kingdom). Centriprep and Centricon centrifugal filter devices were purchased from Millipore Corporation (Billerica, Massachusetts).

E. coli strains CJ236 and JM109 were from Bio-Rad Laboratories GmbH (Munich, Germany). Helper phage M13K07 was from GE Healthcare Bio-Sciences Ltd (United Kingdom).

Restriction enzymes, T4 DNA ligase, T4 DNA polymerase, and polynucleotide kinase were obtained from New England Biolabs (Schwalbach, Germany). Adenosine triphosphate and deoxynucleoside triphosphates were purchased from Roche (Mannheim, Germany). Catalase from bovine liver was obtained from Sigma (St. Louis, MO). Synthesis of oligonucleotides for site-directed mutagenesis and DNA sequencing were performed by MWG Biotech AG (Ebersberg, Germany).

Construction of *E. coli* PDFwt expression vector

Wild-type peptide defomylase from *E. coli* ($EcPDF_{wt}$) was produced with the expression vector pBAD/*Myc*-His-DEST using Gateway cloning technology (Invitrogen). Vector pBAD/*Myc*-His-DEST is a derivative of plasmid pBAD/*Myc*-HisC [31] and has been constructed by insertion of an *attR1-cat*/*ccdB-attR2* cassette at the position of the Shine-Dalgarno sequence in pBAD/*Myc*-HisC, which is located downstream of the strong and tightly controllable *araBAD* promoter 25 [10]. The PDF gene was first amplified by PCR using forward primer 5'-GGGG*ACAAGTTTGTACAAAAAAGCAGGCT*AG-GAGGAATTAACCA**ATG**TCCGTGCTTCAAGTGTTACATATTCC – 3' as (*attB1* site in italics, parts complementary to the *def* gene underlined, start codon in boldface), and reverse primer 5'-GGGG*ACCACTTTGTACAAGAAAGCTGGGT***TTA**AGCCCGGGCTTT -CAGACGATCCAGTTTTTC–3' as (*attB2* site in italics, parts complementary to the *def* gene underlined, stop codon in boldface) and plasmid pTL7-1. The PCR reaction was performed using AccuPrime Taq DNA polymerase (1U) and accompanying buffer (Invitrogen). The amplification reaction was started with an initial denaturation of 2 min at 95°C, followed by 30 cycles of 15 s at 95°C, 30 s at 58°C and 1 min at 68°C, with an additional cycle of 5 min at 68°C. The amplification product was purified using the QIAquick PCR purification kit (Qiagen), after which this product was introduced into the pDONR201 vector via a Gateway BP recombination reaction in a vector:fragment molar ratio of 1:2. After transformation of *E. coli* TOP 10, recombinant cells were selected by

plating on LB plates containing kanamycin (50 µg/mL), followed by overnight incubation at 28°C. The PDF-encoding gene in pDONR201 was subsequently recombined to expression vector pBAD/*Myc*-His-DEST in a standard LR recombination reaction with a molar ratio of destination vector vs. entry vector of 1:2.4 (150:300 ng).

After transformation of *E. coli* TOP10 recombinant cells were selected by plating on LB plates containing carbenicillin (100 µg/mL) followed by overnight incubation at 28°C. Finally, a correct clone, as established by plasmid DNA isolation and restriction enzyme analysis was designated pBAD/Myc-His-DEST PDF_{wt}.

Construction of EcPDF variant expression vector

The method we used for site-directed mutagenesis of the *def* gene is based on the procedure described by Kunkel [24] using uracil-containing ssDNA. It is also described in the instruction manual of the Muta-Gene phagemid in vitro mutagenesis kit by Bio-Rad (Bio-Rad Laboratories GmbH, Munich, Germany).

The *def* gene from *E. coli* K12 (EMBL Nucleotide Sequence Database accession number U00096, nucleotides 343,1712 to 343,2221) was cloned between the *Eco*RI and *Hin*dIII sites of phagemid pTZ18U [24] putting the gene under the control of the *lac* promoter [10]. The resulting phagemid was named p1074 and was used to prepare ssDNA in which some thymidine residues were replaced by uracil. This was done by superinfection of *E. coli* strain CJ236 (*ung*$^-$, *dut*$^-$) harboring plasmid p1074 with helper phage M13K07. *E. coli* strain CJ236 harboring p1074 was grown in 2xYT medium (16 g/L Bacto-Tryptone, 10 g/L yeast extract, 5 g/L NaCl, 10 mM KH_2PO_4 pH 7.5) supplemented with 30 µg/mL chloramphenicol, 150 µg/mL ampicillin and 1 mM thiamine hydrochloride at 37°C with gently shaking (150 rpm). Then 0,5 mL of the culture (OD_{550nm} = 5) were diluted in 100 mL of the same medium and infected with M13K07 At an MOI of 10 (MOI = phages per mL/cells per mL). Infected cells were grown at 37°C and 150 rpm for 20 h (final OD OD_{550nm} = 2.9). After removal of cells by centrifugation (17,000g, 20 min) the uracil-containing ssDNA of p1074 was isolated from the supernatant by precipitation with PEG 6000 (final concentration 5% PEG 6000, 0.625 M NaCl) and purified by extraction with phenol. The yield was 215 µg of pure ssDNA.

The uracil-containing ssDNA of p1074 was used as template for in vitro synthesis of the second strand with the 5'-phosphorylated oligonucleotide as mutagenic primer: as template for the in vitro synthesis of the second strand.

The obtained dsDNA was used to transform *E. coli* JM109 (*ung*$^+$) which contains a proficient uracil *N*-glycosylase that efficiently inactivates the uracil-containing strand of the phagemid leaving the mutated strand for replication. Silent mutations which introduce a new restriction site were incorporated in the mutagenic primers for easy detection of mutants. The mutatations were verified by DNA sequencing of the complete *def* gene.

For the construction of PDF_{Etag} and PDF_{short}, two oligonucleotides were used in separate reactions. These were: prPDF-Etag (5'- CGTCACGGTCCTCTTCTTCTTCCT-CCTCTTCCTCTTCCGAAACATC), which introduces 8 additional codons for Glu between Glu64 and Asn65 and changes Asn65 to Asp, and prPDFstop (GTTGTTT-CACTTAAGACAGATAATCC) which changes the *pdf* Pro148 codon (CCG) into a stop codon (TAA). The oligoglutamate tag is encoded by a mixture of CTC and CTT codons, to prevent hairpin formation. The new restriction sites (*Ava*II in $PDFE_{tag}$ and *Afl*II in PDF_{short}) are underlined. The resulting PDF_{Etag}- and the PDF_{short}-encoding plasmids were respectively called pTL7-1 and pTL7-2.

For expression, the mutated genes were cloned into the pBAD/*Myc*-His-DEST vector as described above. For amplification of $EcPDF_{Etag}$ primer prPDF-for and prPDF-long-rev were used, whereas for the amplification of $EcPDF_{short}$ prPDF-for and prPDF-short-rev were combined. Primer sequences: prPDF–for, GGGGACAAGTTT-GTACAAAAAAGCAGGCTAGGAGGAATTAACCAATGTCCGTGCTTCAAGTGTTACAT-ATTCC; prPDF-long-rev, GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGCCC-GGGCTTTCAGACGATCCAGTTTTTC; prPDF-short-rev, GGGGCCACTTTGTACAA-GAAAGCTGGGTTTAAGACAGATAATCCATAAACAGTTTGCCG. Plasmids pTL7-1 and pTL7-2 were used as template in the PCR reactions, respectively, and the resulting constructs were called pBAD-PDF_{Etag} and pBAD-PDF_{short}.

Cultivation and cell free extract preparation

$EcPDF_{wt}$, $EcPDF_{Etag}$ and $EcPDF_{short}$ were isolated from overproducing *E. coli* TOP10 cells transformed with the respective plasmids. Cells were grown at 28°C for 14-16 h in 1.5 L LB medium containing 100 µg/mL carbenicillin. When the OD_{620} reached 0.6, PDF expression was induced by addition of 0.02 % L-arabinose. Cells were harvested by centrifugation and about 12 g (wet weight) of cell paste was suspended in 40 ml buffer (20 mM HEPES/KOH, 100 mM KCl, pH 7.7, supplemented with 10 µg/mL

catalase from bovine liver). After disintegration by sonication (Branson B12, 10 min) at 0°C, cell-free extract (CFE) was obtained by centrifugation at 33,300 x *g* for 1 h at 4°C.

Purification of PDF using anion exchange chromatography

After preparation of CFE with $EcPDF_{wt}$ and $EcPDF_{Etag}$, 1 mL of the clear extract (5 mg/mL of total protein for PDF_{Etag} and 7.8 mg/mL for PDF_{wt}) was fractionated using a 1 mL Mono Q 5/50 GLcolumn that was equilibrated with a buffer solution composed of 20 mM HEPES/KOH, 1 mM TCEP, 100 mM KCl, pH 7.7. After washing with 15 mL of the same buffer to remove the unbound proteins, PDF was eluted with 60 mL of 20 mM HEPES/KOH, 1 mM TCEP, 1 M KCl, pH 7.7. The fractions with PDF activity were detected using the HPLC method described below and concentrated using Centriprep 30 kDa filters (Millipore) to 10 mg/mL of protein.

This purification procedure was also scaled up. An amount of 44 mL of CFE (30 mg/mL total protein) was loaded on a HiLoad 26/10 Q Sepharose HP column after equilibrating the column with 20 mM HEPES/KOH, 1mM TCEP, 100 mM KCl, pH 7.7. Unbound proteins were washed off with 100 mL of the same buffer solution and the PDF was eluted with 600 mL of 20 mM HEPES/KOH, 1 mM TCEP, 1 M KCl, pH 7.7. The fractions with PDF activity were concentrated using Centriprep 30 kDa filters to 20 mg/mL of protein.

Purification of PDF using affinity chromatography

For purification by affinity column chromatography $EcPDF_{wt}$ CFE was prepared as described above but the cell paste was resuspended in a buffer containing KF as a salt. The buffer composition was 20 mM HEPES/KOH pH 7.7, 100 mM KF, supplemented with 10 µg/mL catalase from bovine liver.

An amount of 10 ml of the clear extract (6.4 mg/mL of total protein) was fractionated using a 20 mL Met-Lys-Sepharose column that had been equilibrated with buffer solution composed of 20 mM HEPES/KOH, 100 mM KF, 0.2 mM TCEP, pH 7.7 [32]. After washing with 42 mL of the same buffer to remove unbound proteins, the PDF was eluted with 40 mL of 20 mM HEPES/KOH, 100 mM KCl, and 0.2 mM TCEP, pH 7.7. The fractions with PDF activity were concentrated by ultrafiltration using Centriprep 10 kDa filters to 10 mg/mL.

Purification of PDF using ultrafiltration

E. coli TOP10 cells containing pBAD-PDF_{wt} were cultivated as described above and about 4 g (wet weight) of cell paste was suspended in 20 mL buffer (20 mM HEPES/KOH, 100 mM KCl, pH 7.7 supplemented with 10 µg/mL catalase from bovine liver). After disintegration by sonication at 0°C, cell-free extract was obtained by centrifugation at 33,300 x *g* for 1 hour. Partial purification of $EcPDF_{wt}$ was performed using Centriprep centrifugal filter units with an Ultracel YM-50 or YM-100 membrane (Millipore). The nominal cut-off of these filters is 50 and 100 kDa, respectively. Before loading the CFE (protein concentration 15 mg/mL) to the Centriprep filter device, the sample was centrifuged for 2 h at 16,100 x *g* and 4^{o}C, the supernatant was collected and subsequently diluted 5-fold with buffer composed of 20 mM HEPES/KOH, 100 mM KCl, supplemented with 10 µg/mL catalase and 2 mM TCEP, pH 7.7. The diluted material (3 mg/mL) was then loaded on a Centriprep YM-50 or YM-100 filter and centrifuged for 2.5 h at 1,000 x *g* and 4°C. When using Centricon YM-30, YM-50 or YM-100 devices the CFE was prepared following the same protocol but centrifugation step was performed for up to 2 h at 1,000 x *g* for the YM-100 and up to 2 h at 3,500 x *g* for the YM-30 and 50.

Purification of PDF using gel filtration chromatography

Cell-free extracts containing $EcPDF_{wt}$, $EcPDF_{Etag}$ or $EcPDF_{short}$ were produced according to the standard protocol. Theextracts had protein concentrations of 10 mg/mL, 8 mg/mL and 12 mg/mL, respectively. The samples were diluted with standard dilution buffer (SDB), which was composed of 20 mM MOPS/NaOH, 100 mM NaCl, supplemented with 10 µg/mL catalase from bovine liver and 2 mM TCEP, pH 7.7, to a protein concentration of 5 mg/ml. Subsequently 100 µL samples of the extracts were chromatographed on a TSK-GEL $G2000SW_{XL}$ column (TOSOH Bioscience). The flow rate used during all the purification steps was 0.5 mL/min.

Enzyme and protein assays

Deformylase activity of PDF was routinely measured at 30°C in a total volume of 500 µL. Samples of 50 µL of different dilutions of PDF in SDB, were mixed with 450 µL of substrate solution composed of 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase and 5.5 mM of the substrate *N*-formyl-Met-Ala-OH. The dilutions of the enzyme were chosen such that approx. 10% of the substrate was

converted in 15 min. After starting the reaction by the addition of the PDF, 100 µL aliquots were withdrawn from the reaction mixture every 5 min and added to 100 µL of 1M phosphate buffer (H_3PO_4-NaOH, pH 2.66) to stop the reaction. Analysis of substrate and product was performed using the HPLC method described below. To obtain an enzymatic blank, 500 µL of the same phosphate stopping buffer and 50 µL of diluted PDF were mixed together prior to the addition of 450 µL of the substrate solution given above.

Peptidase activity in samples was determined using a mixture of the following 8 dipeptides: H-Leu-Phe-OH, H-Phe-Leu-OH, H-Asp-Phe-OH, H-Phe-Asp-OH, H-Arg-Phe-OH, H-Phe-Arg-OH, H-Gly-Phe-OH, H-Phe-Gly-OH. An amount of 450 µL of a solution composed of 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase, and 0.625 mM of each of the 8 dipeptides, was mixed with 50 µL of different dilutions of PDF in SDB. The dilution was chosen such that approx. 10% of the substrate was converted in 15 min. After starting the reaction at 30°C by the addition of the PDF, 50 µL samples were withdrawn from the incubation mixture every 5 min and added to 150 µl methanol to stop the reaction. For the enzymatic blank, the methanol was added prior to the enzyme. Analysis of substrate and product was performed using the HPLC method described below.

Protein concentrations were determined using the Bradford method with bovine serum albumin (BSA, Sigma) as standard protein [34].

PDF stability in the presence of cosolvents

The effect on enzyme stability of several water miscible organic cosolvents [methanol, tetrahydrofuran (THF), dimethylformamide (DMF), tert-butanol, dimethyl sulfoxide (DMSO) and N-methyl-2-pyrrolidone (NMP)] in a 10, 20 or 40% (v/v) concentration was measured by measuring the activity of diluted purified $EcPDF_{wt}$ and its variants at 30°C in a solution containing 20 mM MOPS/NaOH, pH 7.7, 100 mM NaCl, 10 µg/mL bovine liver catalase, and 2 mM TCEP. To this mixture was added 450 µL of substrate solution composed of 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase and 5.5 mM of the substrate *N*-formyl-Met-Ala-OH. Samples were periodically withdrawn from the incubation mixture to measure the specific activity. The residual activity is reported as a percentage of the specific activity of the enzyme that was found when no cosolvent was added to the reaction mixture.

Effect of temperature on PDF stability

The temperature stability of the enzymes was analyzed by preincubating purified PDFs (30 µg/mL) for 60 min at temperatures between 4 and 60 °C in standard dilution buffer. After an incubation of 10 min on ice, 50 µL samples of the pretreated enzyme solutions were used to determine the remaining activity according to the standard assay at 30 °C.

HPLC analyses

HPLC was carried out using a stainless-steel analytical column (250 mm length, 4.6 mm ID) packed with Inertsil ODS-3 material, 5 µm particle size from Alltech Applied Science (Breda, The Netherlands). The flow rate was 1 mL/min. Detection was performed at 40°C at a wavelength of 210 nm. The injection volume was 5 µL.

For-Met-Ala-OH and H-Met-Ala-OH were analyzed using the following gradient of acetonitrile in 10 mM H_3PO_4: t=0-4 min, 0.1 % (v/v) acetonitrile isocratic; t=4-15 min, 0.1%-50% acetonitrile linear increase; t=15-15.1 min, 50%-0.1% linear decrease; t=15.1-20 min, 0.1% acetonitrile isocratic (retention times: For-Met-Ala-OH = 13.54 min, Met-Ala-OH = 5.39 min).

The peptides H-Leu-Phe-NH_2, N-formyl-Leu-Phe-NH_2, and *N*-formyl-Tyr-Leu-Phe-NH_2 were analyzed using the following gradient of acetonitrile in 10 mM H_3PO_4: t=0-5 min, 2.5% (v/v) acetonitrile isocratic; t=5-20 min, 2.5%-75% acetonitrile linear increase; t=20-20.1 min, 75%-2.5% linear decrease; t=20.1-27 min, 2.5% acetonitrile isocratic (retention times: H-Leu-Phe-NH_2 = 12 min, N-formyl-Leu-Phe-NH_2 = 16.81 min, *N*-formyl-Tyr-Leu-Phe-NH_2 = 19.46 min.

For peptidase activity measurements, the production of phenylalanine was analyzed using the following gradient of acetonitrile in 10 mM H_3PO_4: t=0-15 min, 0-30% acetonitrile linear gradient; t=15-15.1 min, 30-0% acetonitrile decrease; t=15.1-20 min, 0% acetonitrile isocratic.

Protection and synthesis of amino acids and peptides

Amino acids, amino acid derivatives, and peptides were formylated using a refluxing mixture of formic acid and 1 equivalent of acetic acid anhydride.

For-Leu-Phe-NH_2 was synthesized as follows. 7.9 g *N*-formyl-Leu-OH and 10 g H-Phe-NH_2 were dissolved in 200 mL of water containing 0.4 mM NaCl and 15 mM $CaCl_2$ keeping the pH at 6 by addition of NaOH (20% w/v). Then, 4 g of thermolysin

(Sigma) was added and the mixture was stirred for 30 h at room temperature. The precipitate was isolated by filtration and washed with 50 mL of water. The identity and purity (>96%) for For-Leu-Phe-NH_2 were confirmed by HPLC analysis

PDF-catalyzed deprotection of For-Leu-Phe-NH_2 was performed as follows. To a solution of 55 mM *N*-formyl-Leu-Phe-NH_2, synthesized as described above, in 260 mM aqueous MOPS-NaOH buffer containing 675 mM NaCl and 0.1 g/L of catalase (pH 7.2) was added purified PDF_{Etag} (obtained using ion-exchange chromatography) to a final concentration of 5 µM. The reaction mixture was incubated at 28°C for 46 h. Product was isolated by extraction with ethyl acetate/H_2O at pH 9.5 and purity was checked by HPLC (>98%).

For the synthesis of For-Tyr-Leu-Phe-NH_2, a solution of 400 mM *N*-formyl-Tyr-OH, 350 mM H-Leu-Phe-NH_2 and 0.1 g of thermolysin was prepared in 4 mL of water containing 0.4 mM NaCl and 15 mM $CaCl_2$. The mixture was incubated at room temperature for 3 days. Product was isolated by filtration and washed first with 10 mL water followed by washing with 5 mL diethyl ether. The identity and purity (>65%) were checked by HPLC.

Results and discussion

It is virtually impossible to predict in a rational manner which separation and purification protocol will be optimal for the isolation of a certain protein. This is due to a lack of knowledge on the molecular properties that determine the exact behavior of the proteins to be separated. A range of purification techniques is available such as fractional precipitation, anion- and cation-exchange chromatography, hydrophobic interaction chromatography, affinity chromatography, gel filtration, and ultra- and microfiltration. Evidently, when being confronted with a complex mixture of proteins, there are many different steps and routes one can take to separate a target protein from components that need to be removed [3]. Optimization of a protocol therefore is very much a matter of trial-and-error, which may make research and development costly. For $EcPDF_{wt}$ and its variants, the major restraints for the choice of a purification strategy are process costs when operated on large scale and industrial applicability. For this reason two methods were investigated: ion-exchange chromatography and ultrafiltration.

Purification of *E. coli* PDF via affinity chromatography

To obtain an expression construct, the *E. coli def* genes coding for $EcPDF_{wt}$ or variants thereof were cloned into pBAD/*Myc*-His-DEST via standard molecular biology procedures. This resulted in a 1,000-2,000 fold overexpression in *E. coli*. $EcPDF_{wt, Etag, short}$ were present in soluble form (Fig. 1). The specific activity of PDF_{wt} in cell-free extract was 229 U/mg as measured with the standard assay, compared to 16 U/mg in a non-overexpressed cell-free extract. The PDF proteins were produced in *E. coli* at levels of 10-20 % of the total soluble protein in CFEs.

For purification of $EcPDF_{wt}$, we first tested the single-step procedure using a Met-Lys-Sepharose affinity column as described in Materials and Methods. In the Met-Lys-Sepharose matrix, the ligand is bound via the ε-NH_2 group of the Lys side-chain to the *N*-hydroxysuccinimide-activated Sepharose. The affinity chromatography method is based on the binding of active PDF to the Met-Lys-Sepharose affinity matrix in the presence of fluoride ions, which significantly increase the affinity of $EcPDF_{wt}$ for small formylated peptides, most likely because fluoride mimics formate [32].

The result of a typical purification experiment starting from ca. 15 g (wet weight) *E. coli* TOP10 cells containing pBAD- PDF_{wt} is given in Table 1. The specific activity of the purified PDF toward the substrate formyl-Met-Ala-OH was 1,880 U/mg protein at pH 7.2 and 30°C. This is somewhat higher than the 1,175 U/mg reported for a homogeneous preparation of PDF(Fe) [32 20], indicating that this approach furnishes pure PDF.

Because it has been shown that the purity of the PDF obtained by affinity chromatography allows its use as a catalyst to deformylate *N*-formyl peptides without concurrent hydrolysis of internal peptide bonds, this method was used here as a reference when measuring the amidase and peptidase content of various enzyme samples. To enable rapid determination of amidase and peptidase activity, enzyme samples were incubated with a mixture of different dipeptides, all containing one Phe residue at the N- or at the C- terminal position. From the activity data obtained the increase of purity was determined by calculating the ratio of the amidase and peptidase activity to the PDF activity for the CFE, and dividing this ratio by the deformylase/amidase ratio found with purified PDF. This calculation showed that in PDF samples isolated by affinity chromatography the level of contamination by amidases was reduced by 1,045-fold.

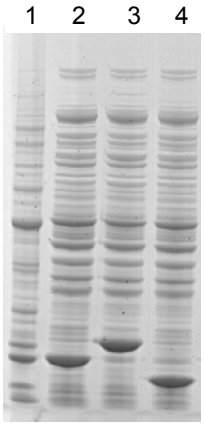


Figure 1. SDS-PAGE of induced *E. coli* cells overexpressing EcPDF.
Lane 1, reference protein ladder;
Lane 2, CFE of *E. coli* cells overexpressing $EcPDF_{wt}$;
Lane 3, CFE of *E. coli* cells overexpressing $EcPDF_{Etag}$;
Lane 4, CFE of *E. coli* cells overexpressing $EcPDF_{short}$.

Besides a relatively low yield of only 57%, the main disadvantage of this purification method is the high cost of the affinity matrix. In fact, the modified Sepharose is not commercially available and its preparation involves a complex chemo-enzymatic process. Moreover, the use of fluoride makes it necessary to take special precautions in order to avoid corrosion of the equipment when applied on industrial scale. Therefore we explored the use of alternative large-scale purification procedures.

Table 1. Purification of $EcPDF_{wt}$ from *E. coli* TOP10 using Met-Lys-Sepharose affinity chromatography.

Purification step	Activity (x 10^3 U)[a]	Protein (mg)	Sp. Act. (U/mg)[a]	Recovery (%)	Purification (fold)	Sp. Act. Amidases (U/mg)	Sp. $Act._{PDF}$/ Sp. $Act._{amidase}$
CFE	14.7	64.5	229	100	1	0.23	995
Met-Lys-Sepharose column	8.4	4.5	1,880	57	8.2	0.002	$1*10^6$

[a] Obtained using For-Met-Ala as substrate.

Purification of PDF via ultrafiltration

Data on the use of ultrafiltration (UF) for protein purification purposes are limited. Most of the published applications of ultrafiltration for protein isolation concern the separation of artificial mixtures. While this has led to a better understanding of the mechanism of protein behavior and separation, lack of application based research has kept ultrafiltration in the blind spot of potential users [17]. Despite this, ultrafiltration could offer a simple and cost effective way of separating PDF from amidases and peptidases present in CFEs. UF processes are cost effective, easy to scale-up and can be fine-tuned to achieve high productivity and product purity at the same time [18].

The size of EcPDF, a monomeric globular protein of about 19 kDa [6] is rather small compared to the size of most proteins in *E. coli*. Considering that the most important amidases and peptidases present in this bacterium are multimers of monomers with a size of about 60 kDa, the selection of an appropriate membrane cut-off could allow their separation from peptide deformylase [1]. To favor this process even more by increasing the differences in enzyme sizes, the use of a smaller version of the EcPDF was also investigated. This smaller version ($EcPDF_{short}$) was readily obtained by introduction of a stop codon in the gene as codon 148, thereby removing residues 148 to 169 from the C-terminal α-helix of the EcPDF and providing an enzyme of 16.7 kDa. It is known that the last 18 residues of EcPDF are disordered in the X-ray structure and dispensable for activity [25].

Based on the native molecular mass of the EcPDF and the shorter variant, two different membrane sizes were tested for PDF retention. Initially $EcPDF_{wt}$ and $EcPDF_{short}$ were loaded on 30 and 50 kDa Centricon filters. For both enzymes a high resistance and very low flux was observed and no PDF activity was detected in the filtrate. Visualization of the 50 kDa filtrate on SDS-PAGE showed the presence of only a very small amount of protein at the position where PDF was expected in view of its molecular weight. Further analysis of mass balance and activity showed that all PDF was still present in the retentate (Fig. 2). Therefore, we decided to use a higher membrane cut-off value. $EcPDF_{wt}$ and $EcPDF_{short}$ were loaded on a 100 kDa Centricon membrane and in this case about 10% of the PDF activity was found in the filtrate. Since no significant differences were observed between yields obtained with $EcPDF_{wt}$ and its shorter variant, further tests with ultrafiltration were carried out only using $EcPDF_{wt}$.

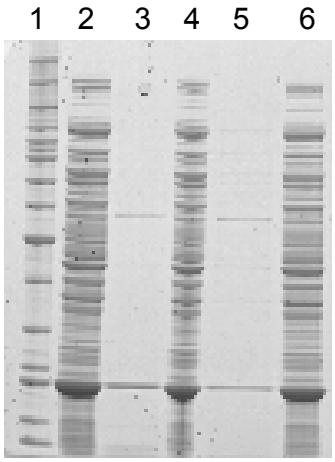


Figure 2. SDS-PAGE of retentate, filtrate and CFE using 30 and 50 kDa Centricon filters.
Lane 1, reference protein ladder;
Lane 2, retentate using 30 kDa Centricon filters;
Lane 3, filtrate using 30 kDa Centricon filters;
Lane 4, retentate using 50 kDa Centricon filters;
Lane 5, filtrate using 50 kDa Centricon filters;
Lane 6, CFE of *E. coli* cells overexpressing $EcPDF_{wt}$.

First, $EcPDF_{wt}$ (4 mg/mL) that had been purified by affinity chromatography was loaded on 50 and 100 kDa Centricon filters. This experiment was carried on to evaluate the capability of the pure enzyme to pass through the filters and the possibility that sedimentation of particles during filtration caused membrane fouling. In the case of the 100 kDa filters, all deformylation activity appeared in the filtrate, whereas the use of 50 kDa cutoff membranes led to only 2% yield of PDF in the filtrate. This was improved to 25% yield when the enzyme was loaded on the 50 kDa filter after 5-fold dilution. Thus passage of PDF_{wt} through the filter is heavily influenced both by the size of the pores in the membrane and protein concentration.

To avoid membrane clogging due to sedimentation of particles, Centriprep filters were used for subsequent experiments. These filters have a design that allows filtration to occur while at the same time fine particles are sedimenting at the bottom of the filter instead of on the membrane. Initially, with 50 kDa Centriprep filters and CFE of *E. coli* overexpressing $EcPDF_{wt}$, no peptide deformylase activity was found in the filtrate. Therefore the same test experiment was performed using a 100 kDa membrane, which resulted in recovery of 21% of the PDF activity in the filtrate (Table 2). The level of contaminating amidase in the purified sample decreased 27-fold compared to CFE.

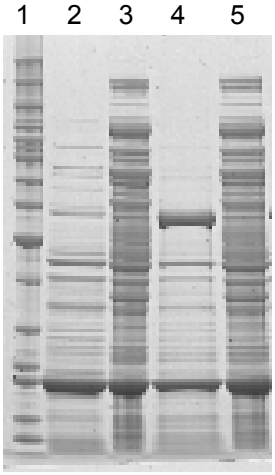


Figure 3. SDS-PAGE of retentate, filtrate and CFE obtained with 100 kDa Centriprep filters.
Lane 1, reference proteins;
Lane 2, filtrate using 100 kDa filters;
Lane 3, retentate using 100 kDa filters;
Lane 4, filtrate using 100 kDa filters;
Lane 5, retentate using 100 kDa filters.

When compared to the results obtained using affinity chromatography (Table 1) ultrafiltration appeared an inefficient purification method. PDF recovery was lower and the amount of amidases remaining in the sample was much higher with ultrafiltration than with affinity chromatography.

The poor performance of ultrafiltration could possibly be due to the formation of protein aggregates that reduce PDF recovery in the filtrate. Different conditions may contribute to formation of aggregates, including protein concentration, pH, ionic strength, and mechanical stress. During ultrafiltration, the protein concentration in the retentate increases, which may cause aggregation and membrane fouling. Moreover, the continuous mechanical stress can cause perturbations of the protein structure with subsequent exposure of hydrophobic surfaces leading to aggregation [12]. The improved yield obtained when diluting the enzyme five-fold in dilution buffer before loading it on the 50 kDa Centricon filter supported this explanation.

One of the methods to prevent the formation of aggregates is the use of additives like salt (0.1 and 1M NaCl) or betaine. However, when this was attempted, no increased yield was observed compared to when standard buffer was used. Membrane fouling is strongly affected by the shear flow causing absorption and/or gel-like depositions on the membrane surface, which may be reduced by cross-flow filtration. Stamatakis and Tien [33] reported that with this technique only a fraction of the particles brought to the membrane surface could deposit on the membrane itself. To test cross-flow filtration, we loaded CFE containing overexpressed $EcPDF_{wt}$ on a 100 kDa cutoff filter. One liter of cell-free extract containing overexpressed EcPDFwt was worked-up as described in

Materials and Methods and subsequently loaded onto the filter. A pressure of 1.5 bars was applied and the retentate was washed three times with process water or MOPS buffer to prevent protein aggregation. Despite this, the amount of PDF that appeared in the flow-through was only 2%.

Table 2. Purification of $EcPDF_{wt}$ from *E. coli* TOP10 using 100 kDa Centriprep filters

Purification step	Activity (x 10^3 U)[a]	Protein (mg)	Sp. Act. (U/mg)[a]	Recovery (%)	Purification (fold)	Sp. Act. amidases (U/mg)	Act.PDF/ Act. amidase
CFE	1.9	10.2	195	100	1	0.2	975
Filtrate	0.4	1.3	321	21	1.6	0.012	$2.7*10^4$

[a] Measured with For-Met-Ala-OH as substrate.

A possible explanation for the unsatisfactory result of cross-flow filtration is that also under these conditions membrane fouling by protein can still occur, depending on the nature of the filtered material. An example of this is reported in the literature by Bailey and Meagher [5]. In their studies concerning the use of cross-flow filtration for clarification of *E. coli* lysates, fouling occurred with some of the membrane materials, leading to poor flux and transmission, despite the use of shear flow and additives [4;5].

In the case of EcPDF we can also speculate that the fouling and clogging of the membranes may be caused by a more specific protein-protein interaction. Bingel-Erlenmeyer and coworkers provided biochemical evidence that EcPDF interacts directly with the ribosomes via its C-terminal α-helix [9]. One *E. coli* cell contains about 15,000 ribosomes, representing 25% of the total mass of the bacterial cell. The molecular weight of ribosomes is about 3,000 kDa. Formation of aggregates of EcPDF and ribosomes or with other proteins could therefore partially prevent its purification via ultrafiltration even when using high molecular weight membranes. Since the preparation of CFE included a centrifugation step (33,300 x g, 60 min), we assume that the lack of passage through ultrafiltration membranes is not the result of the presence of the active PDF in inclusion bodies, as these would end up in the pellet fraction.

Purification of PDF via ion-exchange chromatography

Ion-exchange chromatography (IEC) represents a cheap and easily scalable protein purification method. For industrial application the use of a single step purification is preferred, therefore $EcPDF_{wt}$ was purified via an optimized gradient elution IEC as described in Materials and Methods. The results of a typical purification experiment starting from 1 mL of the clear cell free extract are given in Table 3.

Table 3. Purification of $EcPDF_{wt}$ from *E. coli* TOP10 containing pBAD-PDF_{wt} using IEC.

Purification step	Activity (x 10^3 U)[a]	Protein (mg)	Sp. Act. (U/mg)[a]	Recovery (%)	Purification (fold)	Sp. Act. amidases (U/mg)	Act.PDF/ Act. amidase
CFE	1.6	7.8	208	100	1	0.2	1040
IEC column	0.3	1.6	180	19	0.9	0.08	2250

[a] Activity measured with For-Met-Ala-OH as substrate.

The specific activity of the purified PDF_{wt} as well as the recovery obtained using this procedure were unsatisfactory since the yield was only 19%. Moreover, the level of contamination due to the presence of amidases in the sample was only reduced by a factor of 2, even when the fractions containing the highest levels of PDF activity were pooled and tested.

These disappointing results of IEC chromatography did not point to clear possibilities for improvement. Therefore, the use of protein engineering to improve the purification procedure was attempted. Recombinant DNA technology allows the fusion of purification tags to the target protein to alter its properties and facilitate the separation by IEC. For example, it has been reported that a polyarginine tag consisting of six arginines fused to the C-terminal end of human urogastrone can act as an ion-exchange tag [21;29]. Other positively and negatively charged tags composed of multiple arginine, glutamate or aspartate residues have also been constructed [13;19;36-38]. In view of these results, we inserted an octaglutamate tag (Etag) in a flexible loop that is present between β-strands C and D (residues 61-66) of PDF enzyme as described in Materials and Methods. It has been reported that this CD loop has to be flexible but has no role in activity and/or substrate binding. Moreover, it is not conserved and its composition and length varies in different PDF types [16;39].The point of insertion is after codon TCG 192 which encodes Glu-64 in the wild-type sequence. The construct was verified by sequencing and expressed in *E. coli* TOP10 yielding the desired $EcPDF_{Etag}$.

The overexpression of this new PDF variant was similar to what was achieved with $EcPDF_{wt}$ (20% of the total protein) and the activity of the $EcPDF_{Etag}$ was only slightly affected by the presence of the charged moiety. After purification using Met-Lys-Sepharose affinity chromatography, the specific activity of $EcPDF_{wt}$ towards the substrate For-Met-Ala-OH was 1,800 U/mg while with PDF_{Etag} an activity of 1300 U/mg was found.

Table 4. Purification of $EcPDF_{Etag}$ from CFE of *E. coli* TOP10 containing pBAD-PDF_{Etag} using IEC with a Mono Q 5/50 GL column.

Purification step	Activity (x 10^3 U)[a]	Protein (mg)	Sp. Act. (U/mg)[a]	Recovery (%)	Purification (fold)	Sp. Act. amidases (U/mg)	Act.PDF/ Act.amidase
CFE	0.8	4.5	180	100	1	0.16	1,120
IEC column	0.6	0.6	970	73	5	$6*10^{-4}$	$1.6*10^6$

[a] Activity measured with For-Met-Ala-OH as substrate.

The results of an IEC purification experiment starting from 1 mL of CFE containing $EcPDF_{Etag}$ are reported in Table 4. More than 70% of the PDF activity was found back after IEC and the amount of amidases in the purified sample was reduced by a factor of 1,500.

The results show that IEC in combination with the use of an engineered PDF carrying a charged octaglutamate tag is a suitable method for the removal of amidases and peptidases from EcPDF to be used in peptide deformylation.

PDF catalyzed deprotection of *N*-formyl-dipeptides

In order to prepare sufficient $EcPDF_{Etag}$ for testing the application of the Etag modified PDF variant in peptide synthesis, the PDF_{Etag} protein was purified from CFE using a HiLoad 26/10 column. This procedure allowed the isolation, in a single step, of 150 mg of pure $EcPDF_{Etag}$ from only 44 mL cell-free extract (protein content 30 mg/mL). The presence of amidases in the sample was reduced 1,300 fold (Table 5, Fig. 4).

Table 5. Purification of $EcPDF_{Etag}$ from CFE of *E. coli* TOP10 containing pBAD-PDF_{Etag} using a HiLoad 26/10 ion-exchange column.

Purification step	Activity (x 10^3 U)[a]	Protein (mg)	Sp. Act. (U/mg)[a]	Recovery (%)	Purification (fold)	Sp. Act. amidases (U/mg)	Act.PDF/ Act.amidase
CFE	21	1400	156	100	1	0.19	821
HiLoad 26/10 column	14	150	884	64	5.5	$8*10^{-4}$	$1.1*10^6$

[a] Activity measured with For-Met-Ala-OH as substrate.

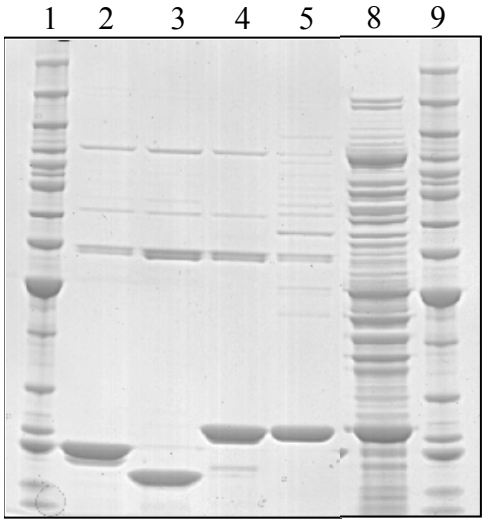


Figure 4. SDS-PAGE of purified EcPDFs variants using IEC and affinity chromatography.
Lane 1, reference proteins
Lane 2, $EcPDF_{wt}$ purified via affinity chromatography;
Lane 3, $EcPDF_{short}$ purified via affinity chromatography;
Lane 4, $EcPDF_{Etag}$ purified via affinity chromatography;
Lane 5, $EcPDF_{Etag}$ purified via IEC;
Lane 8, CFE , $EcPDF_{Etag}$;
Lane 9, reference proteins.

The purified $EcPDF_{Etag}$ was then used in a reaction containing 1.2 mM *N*-formyl-Leu-Phe-NH_2. The mixture was incubated under standard reaction conditions (pH 7.2, 30°C, 8 µg/mL $EcPDF_{Etag}$). After 4 h, full deformylation of the substrate was obtained without any detectable peptide bond hydrolysis. When the deformylation reaction was carried out with CFE, extensive peptide bond hydrolysis occurred, most likely due to leucine aminopeptidase activity produced by the *E. coli* host.

Cosolvent stability ofEcPDFwt and EcPDFEtag

A major problem associated with the application of enzymes for the deprotection of *N*-formyl peptides is the insolubility of protected peptides in water. Polar organic solvents, such as dimethyl sulfoxide (DMSO), dimethylformamide (DMF), methanol (MeOH), *N*-methylpyrrolidone (NMP), tetrahydrofuran (THF) and 2-methyl-2-propanol

are often utilized as cosolvents to increase peptides solubility [27]. Therefore, the stability of PDF_{wt}, PDF_{short} and PDF_{Etag} purified via affinity chromatography was tested by determining activities in the presence of different concentrations of cosolvents. Complete inactivation of these PDFs was observed when the enzymes were incubated and their activity tested with 10% (vol/vol) NMP and THF. The effect of DMF and DMSO was not as detrimental, but remaining activities were very low for all three enzymes, even when only 10% cosolvent was used (Fig. 5).

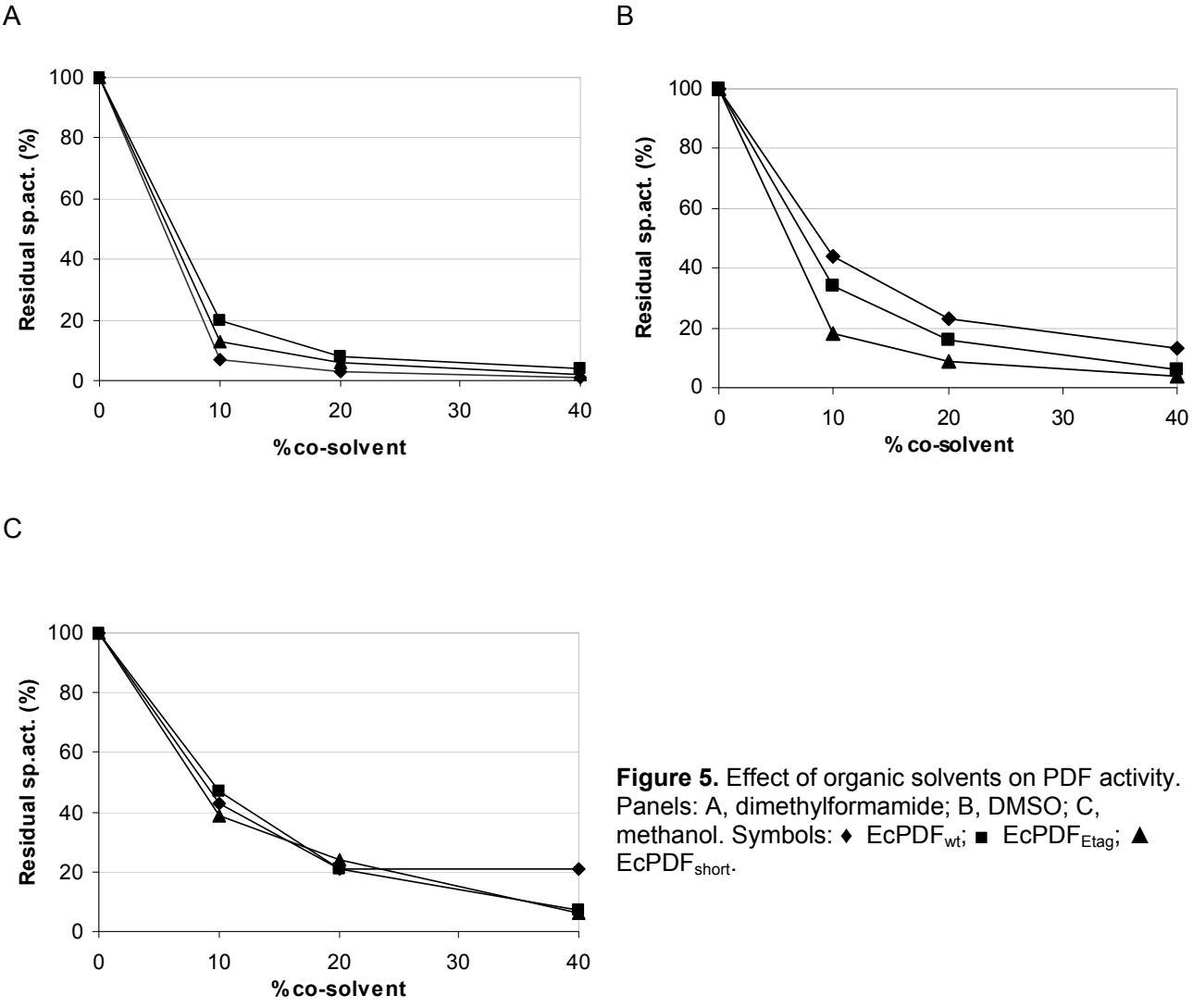


Figure 5. Effect of organic solvents on PDF activity. Panels: A, dimethylformamide; B, DMSO; C, methanol. Symbols: ♦ $EcPDF_{wt}$; ■ $EcPDF_{Etag}$; ▲ $EcPDF_{short}$.

When the same test was carried out by incubating the PDFs in different concentrations of methanol, the loss of activity was lower. Despite this better stability in

methanol, the residual activity was still low; only 20% of the initial activity was left in presence of 20% (vol/vol) methanol (Fig. 5).

The data obtained show that there is only little difference in enzyme stability between the $EcPDF_{wt}$ and the two engineered variants. Apparently, neither the insertion of the negatively charged polyglutamate tag in the flexible loop nor the removal of the C-terminal α-helix had a major effect on enzyme stability in presence of cosolvents.

Effect of temperature on enzyme stability

The thermal stability of the EcPDF and its variants was determined by measuring the remaining activity after preincubation of the purified enzymes for 60 min at temperatures between 4 and 60 °C. The wild-type PDF appeared to be more stable than the mutants at temperatures up to 50 °C. Nevertheless, also PDF_{wt} suffered from considerable thermal deactivation, e.g. 60% of the initial activity was left after preincubation at 40 °C. Furthermore, at temperatures up to 20 °C there is no difference in activity between the three enzymes (Fig. 6A).

The temperature optimum was measured with For-Met-Ala as the substrate under standard reaction conditions. The maximum activity of the three enzymes appeared to be at 30°C. Despite a major activity loss at higher temperatures, a stabilizing effect due to the presence of substrate was visible for temperatures close to the optimum (Fig. 6).

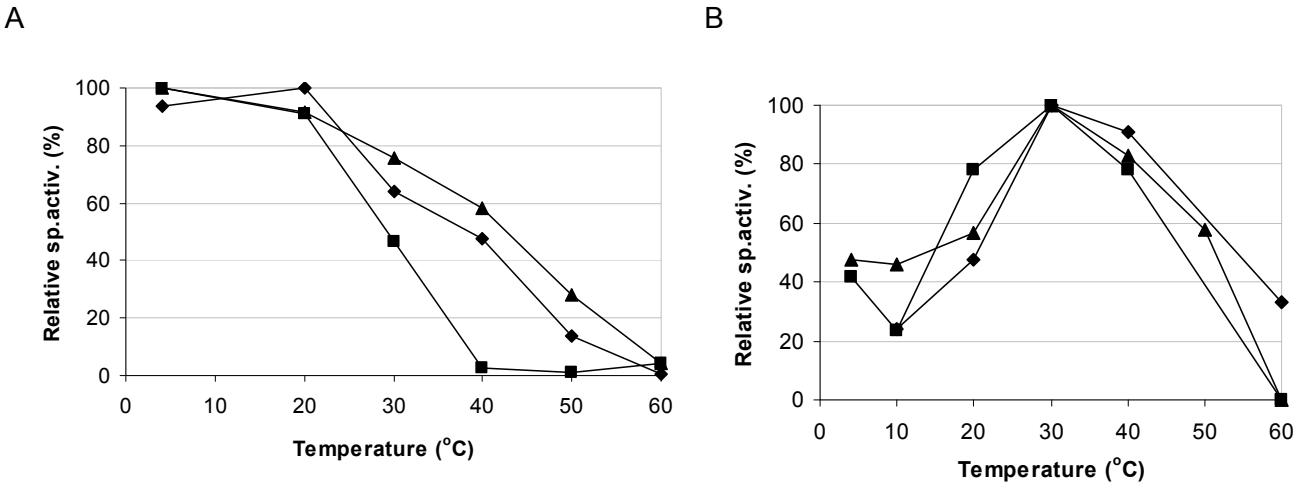


Figure 6. Effect of temperature on PDF. A) Effect of temperature on enzyme stability; B) Temperature optimum with For-Met-Ala-OH as substrate. Symols: ♦, $EcPDF_{wt}$; ■, $EcPDF_{Etag}$; ▲, $EcPDF_{short}$.

The storage stability of the purified enzymes was also monitored, both at -20 °C and 4 °C (Fig. 7). When stored up to 60 days, all PDFs showed a good stability at both storage temperatures. Only after 120 days at 4 °C, a decrease in activity was observed for the $EcPDF_{Etag}$ and $EcPDF_{short}$, while the wild-type enzyme maintained its original activity. Based on these results, all purified PDF samples were stored at -20 °C.

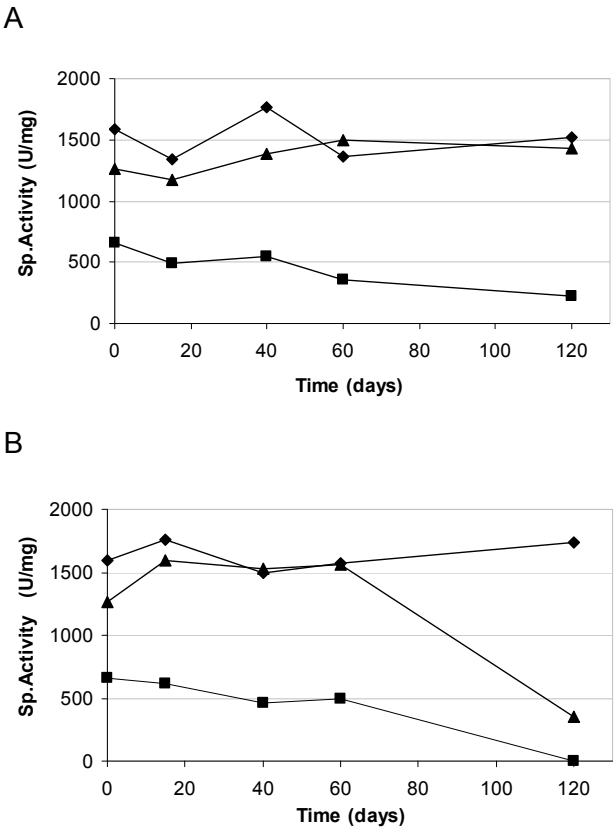


Figure 7. Storage stability at -20°C (panel A) and 4°C (panel B). Symbols:
♦, PDF_{wt};
■, PDF_{Etag};
▲, PDF_{short}.

Synthesis of the tripeptide *N*-formyl-Tyr-Leu-Phe-NH2 using E-tagged peptide deformylase and thermolysin

In order to demonstrate the applicability of the purified $EcPDF_{Etag}$ in chemo-enzymatic peptide synthesis, we synthesized the tripeptide *N*-formyl-Tyr-Leu-Phe-NH_2 using two thermolysin-catalyzed coupling steps and one PDF-catalyzed deformylation step according to the scheme given in Fig. 8. The formyl group can be conveniently introduced on single amino acids and peptides using formic acid and acetic acid anhydride. First, a dipeptide was made by coupling the formyl-protected amino acid *N*-formyl-Leu with H-Phe-NH_2 using thermolysin (Fig. 8, step a). The formylated dipeptide precipitated from the reaction mixture and was isolated. Subsequently, it was

deprotected with $EcPDF_{Etag}$ (Fig. 8, step b). This resulted in complete removal of the formyl group, without any detectable peptide bond hydrolysis. In addition, since PDF is highly L-enantioselective, only the correct enantiomer would undergo deformylation, allowing a high optical purity of the obtained product. After isolation of H-Leu-Phe-NH_2 with ethyl acetate/H_2O at pH 9.5, the purity of the product was >98% with a final yield of 80%. The deprotected Leu-Phe-NH_2 dipeptide was subsequently coupled to *N*-formylated tyrosine applying thermolysin as catalyst (Fig. 8c). This yielded the expected tripeptide *N*-formyl-Tyr-Leu-Phe-NH_2.

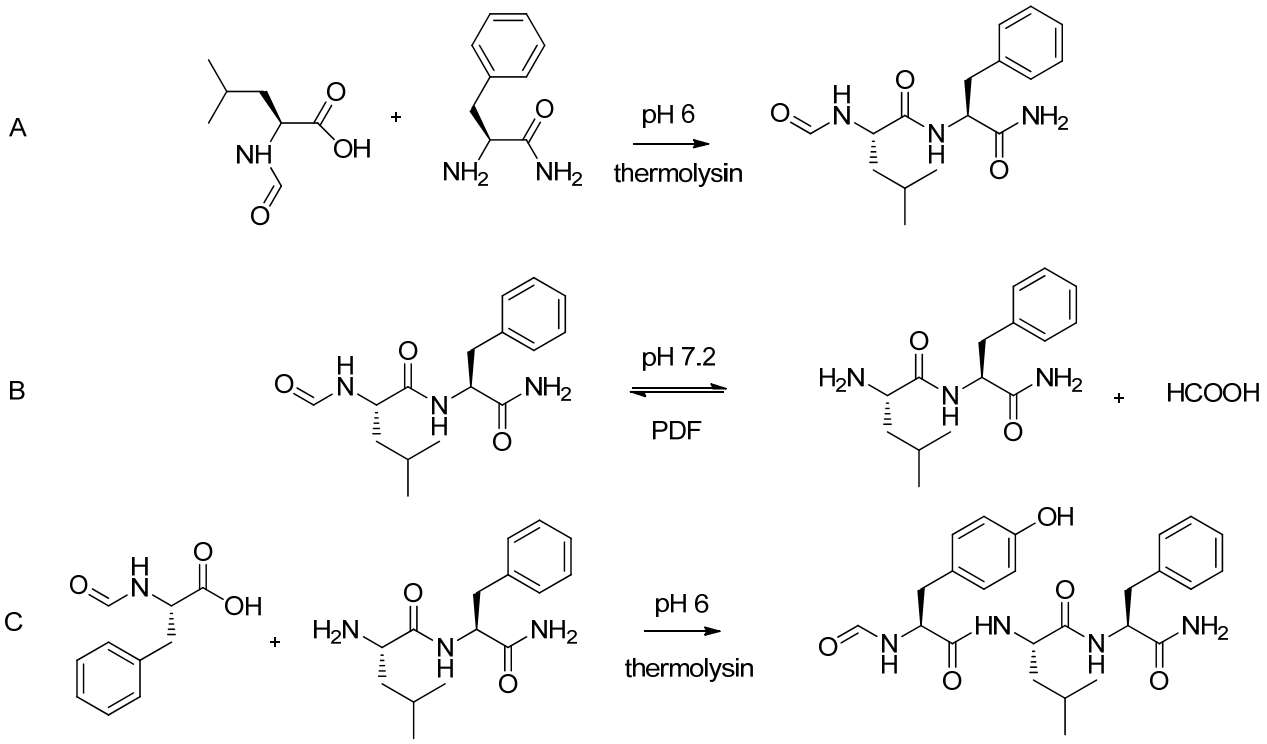


Figure 8. Reaction scheme of the synthesis of *N*-formyl-Tyr-Leu-Phe-NH_2 starting from amino acids and amino acids derivatives as building blocks.

Conclusions

A PDF variant carrying a polyglutamate tag of 9 residues inserted into a flexible surface loop of the enzyme was successfully overexpressed in *E. coli* TOP10. The activity and stability of the enzyme were checked and it was found that the properties of

the engineered variant are comparable to those of the wild-type form, with the exception of a reduced thermostability at temperatures above 20 °C. Similar negatively charged tags composed of glutamate or aspartate groups have been fused to green fluorescent protein, maltose binding protein and firefly luciferase to allow rapid ion-exchange purification [21]. More general purification tags providing different interaction possibilities have also been proposed [7]. Earlier it was described that polyarginine tags can increase the isolelectric point of human β-urogastrone and facilitate isolation [30].

$EcPDF_{Etag}$ was subsequently purified from the overproducing strain by using ion exchange chromatography with a 64% recovery and a purification factor of 5. This method has the advantage of being industrially applicable, cheap and easily scalable. Compared to the wild-type enzyme lacking this tag on the internal flexible loop, the E-tagged protein was more easily purified. After a single chromatographic step, protein of about 80% purity was obtained, lacking most of the amidase activity that was detected in cell-free extract and that is detrimental for peptide synthesis applications. The stability and activity of the E-tagged enzyme were not significantly different from those of wild-type enzyme.

The enzyme appeared to be suitable for the application in peptide synthesis since it allows a mild and selective deprotection of *N*-formyl peptides without significant peptide bond hydrolysis. An additional advantage of the PDF-catalyzed deprotection is that N-terminal *R*-isomers are not deformylated, leading to an increased diastereomeric purity of the deformylated peptide [32].

References

1. **Rao, MB, A,M, Tanksale, M.S. Ghatge, V.V. Deshpande.** 1998. Molecular and biotechnological aspects of microbial proteases. Microbiol. Mol. Biol. Rev. **62**: 597-635.
2. **Adams, J. M.** 1968. On the release of the formyl group from nascent protein. J. Mol. Biol. **33**:571-574.
3. **Asenjo, J. and B. Andrews**. 2008. Protein purification using chromatography: selection of type, modelling and optimization of operating conditions. J. Mol. Recognit. **22**:65-76.
4. **Bailey, S. M. and M. M. Meagher**. 1997. The effect of denaturants on the crossflow membrane filtration of *Escherichia coli* lysates containing inclusion bodies. J. Membr. Science **131**:29-38.
5. **Bailey, S. and M. Meagher**. 1997. Crossflow microfiltration of recombinant *Escherichia coli* lysates after high pressure homogenization. Biotech. Bioeng. **56**:304-310.

6. **Becker, A., I. Schlichting, W. Kabsch, S. Schultz, and A. F. Wagner**. 1998. Structure of peptide deformylase and identification of the substrate binding site. J. Biol. Chem. **273**:11413-11416.
7. **Becker, K., J. Van Alstine, and L. Bnlow**. 2008. Multipurpose peptide tags for protein isolation. J. Chromatogr. A **1202**:40-46.
8. **Ben Bassat, A., K. Bauer, S. Y. Chang, K. Myambo, A. Boosman, and S. Chang**. 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J. Bacteriol. **169**:751-757.
9. **Bingel-Erlenmeyer, R., R. Kohler, G. Kramer, A. Sandikci, S. Antolic, T. Maier, C. Schaffitzel, B. Wiedmann, B. Bukau, and N. Ban**. 2008. A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature **452**:108-111.
10. **Boesten, W.B., P. Raemakers-Franken, T. Sonke, G. Euverink and P. Grijpstra.** 2003. Polypeptides having alpha-H-alpha amino acide amide racemase activity and nucleic acids encoding the same. US Patent 7491520.
11. **Bordusa, F.** 2002. Proteases in organic synthesis. Chem. Rev. **102**:4817-4868.
12. **Chi, E. Y., S. Krishnan, T. W. Randolph, and J. F. Carpenter**. 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm. Res. **20**:1325-1336.
13. **Dalboge, H., H. H. Dahl, J. Pedersen, J. W. Hansen, and T. Christensen**. 1987. A novel enzymatic method for production of authentic hGH from an *Escherichia coli* produced hGH-Precursor. Nat. Biotech. **5**:161-164.
14. **Donnan, S.** 1996. Protein Purification Protocols: Second Ed: Human Press Inc..
15. **Escobar-Alvarez, S., Y. Goldgur, G. Yang, O. Ouerfelli, Y. Li, and D. A. Scheinberg**. 2010. Structure and activity of human mitochondrial peptide deformylase, a novel cancer target. Mol. Cell. Biol. **30:**5099-5109.
16. **Fieulaine, S., C. Juillan-Binard, A. Serero, F. Dardel, C. Giglione, T. Meinnel, and J. L. Ferrer**. 2005. The crystal structure of mitochondrial (Type 1A) peptide deformylase provides clear guidelines for the design of inhibitors specific for the bacterial forms. J. Biol. Chem. **280**:42315-42324.
17. **Ghosh, R. and Z. F. Cui**. 2000. Protein purification by ultrafiltration with pre-treated membrane. J. Membr. Science **167**:47-53.
18. **Ghosh, R., S. S. Silva, and Z. Cui**. 2000. Lysozyme separation by hollow-fibre ultrafiltration. Biochem. Eng. J. **6**:19-24.
19. **Graslund, T., G. Lundin, M. Uhlen, P. A. Nygren, and S. Hober**. 2000. Charge engineering of a protein domain to allow efficient ion-exchange recovery. Prot. Eng., Design Selection **13**:703-709.
20. **Groche, D., A. Becker, I. Schlichting, W. Kabsch, S. Schultz, and A. F. V. Wagner**. 1998. Isolation and crystallization of functionally competent *Escherichia coli* peptide deformylase forms containing either iron or nickel in the active site. Biochem. Biophys. Res. Comm. **246**:342-346.
21. **Hedhammar, M., T. Graslund, M. Uhlen, and S. Hober**. 2004. Negatively charged purification tags for selective anion-exchange recovery. Prot. Eng., Design Selection **17**:779-786.
22. **Leeds, J. A. and C. R. Dean**. 2006. Peptide deformylase as an antibacterial target: a critical assessment. Curr. Opin. Pharmac. **6**:445-452.

23. **Mannervik, B., A. Runarsdottir, and S. Kurtovic**. 2009. Multi-substrate coactivity space and quasi-species in enzyme evolution: Ohno's dilemma, promiscuity and functional orthogonality. Biochem. Society Trans. **037**:740-744.

24. **Mead, D. A., E. Szczesna-Skorupa, and B. Kemper**. 1986. Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Prot. Eng., Design Selection **1**:67-74.

25. **Meinnel, T., C. Lazennec, F. Dardel, J. M. Schmitter, and S. Blanquet**. 1996. The C-terminal domain of peptide deformylase is disordered and dispensable for activity. FEBS Letters **385**:91-95.

26. **Muramatsu, I., M. Murakami, T. Yoneda, and A. Hagitani**. 1964. The formylation of amino acids with acetic formic anhydride. Bull. Chem. Soc. Jpn. **38**:244-246.

27. **Narita, M., S. Honda, H. Umeyama, and S. Obana**. 1988. The solubility of peptide intermediates in organic solvents. Solubilizing potential of hexafluoro-2-propanol. Bull. Chem. Soc. Jpn. **61**:281-284.

28. **Rajagopalan, P. T. R., A. Datta, and D. Pei**. 1997. Purification, characterization, and inhibition of peptide deformylase from *Escherichia coli*. Biochemistry **36**:13910-13918.

29. **Sassenfeld, H. M. and S. J. Brewer**. 1984. A polypeptide fusion designed for the purification of recombinant proteins. Nat. Biotech. **2**:76-81.

30. **Smith, J. C., R. B. Derbyshire, E. Cook, L. Dunthorne, J. Viney, S. J. Brewer, H. M. Sassenfeld, and L. D. Bell**. 1984. Chemical synthesis and cloning of a poly(arginine)-coding gene fragment designed to aid polypeptide purification. Gene **32**:321-327.

31. **Sonke T**. 2008. Novel developments in the chemo-enzymatic synthesis of enantiopure α-hydrogen-and α,α-disubstituted α-amino acids and derivatives. PhD thesis, University of Amsterdam.

32. **Sonke, T., B. Kaptein, A. F. V. Wagner, P. J. L. M. Quaedflieg, S. Schultz, S. Ernste, A. Schepers, J. H. M. Mommers, and Q. B. Broxterman**. 2004. Peptide deformylase as biocatalyst for the synthesis of enantiomerically pure amino acid derivatives. J. Mol. Catal. B: Enzymatic **29**:265-277.

33. **Stamatakis, K., and C. Tien**. 1992. A simple model of cross-flow filtration based on particle adhesion. AIChE Journal **39**:1292-1302.

34. **Stoscheck, C.** 1990. Quantitation of protein. Methods Enzymol. **182**:50-69.

35. **Toyonaka T., F. Kawasaki, and Y and Yokohama Y.** 1978. Method of removing formyl groups from N-formyl-amino acid and N-formyl-peptide esters having free carboxyl groups. US Patent 4071511.

36. **Uhlθn, M., B. r. Nilsson, B. Guss, M. Lindberg, S. Gatenbeck, and L. Philipson**. 1983. Gene fusion vectors based on the gene for staphylococcal protein A. Gene **23**:369-378.

37. **Zhang, C., R. Love, J. Jilka, and C. Glatz**. 2000. Genetic engineering strategies for purification of recombinant proteins from canola by anion exchange chromatography: an example of β-glucuronidase. Biotechnology Progress **17**: 161.

38. **Zhao, J., C. Ford, C. Glatz, M. Rougvie, and S. Gendel**. 1990. Polyelectrolyte precipitation of beta-galactosidase fusions containing poly-aspartic acid tails. J. Biotech. **14**:273-284.

39. **Zhou, Z., X. Song, Y. Li, and W. Gong**. 2004. Unique structural characteristics of peptide deformylase from pathogenic bacterium *Leptospira interrogans*. J. Mol. Biol. **339**:207-215.

Chapter 3

Exploring bacterial genomes for novel PDFs

Abstract

Peptide deformylases are widely studied for their potential as chemotherapeutic targets, but their use in biocatalytic applications has been hardly explored. To find peptide deformylases that show suitable characteristics for industrially relevant biotransformations, peptide deformylase sequences were retrieved from the genomes of a variety of organisms. Using multiple sequence alignments it was found that these enzymes can be classified in 8 phylogenetic groups. To enable determination of the substrate profile, temperature optimum, thermostability, and tolerance to organic solvents, eight bacterial peptide deformylase genes from different phylogenetic groups were cloned and expressed in *Escherichia coli*. The results showed good overexpression of peptide deformylases in active form and revealed that the enzymes had very similar substrate specificities. Temperature optima, on the other hand, varied from 25 to 50°C.

Introduction

In eubacteria, mitochondria and chloroplasts protein synthesis is a complex process that requires the removal of an *N*-formyl group that is present at the *N*-terminus of the nascent polypeptide chain by a peptide deformylase (PDF) [24;26]. This step is crucial for the subsequent removal of methionine by methionine aminopeptidase and the maturation of the polypeptide [14]. For many years, the *N*-terminal deformylation of peptides after translation was considered a unique feature of bacterial cells, but more recently, DNA sequences similar to the genes encoding peptide deformylase (*def*) in bacteria were found in unicellular eukaryotes, plants and mammals [14;23]. The eukaryotic PDFs were proven to be active *in vivo* and *in vitro* but their biological role is still not completely clarified [5;11;28].

The central role of PDF in bacterial protein synthesis triggered efforts to discover antibiotics that can selectively target this enzyme [20]. This is now complicated by the discovery of active human forms of PDF (HsPDF). A key step in the design of inhibitors, which should not be toxic to mammalian cells, is the elucidation of structures of bacterial and human PDFs [11]. Crystal structures of several peptide deformylases are now available and inhibitors can be designed on the basis of structural differences, for example, by targeting the active-site entrance [11]

Despite the vast amount of research on PDFs driven by the desire to develop effective antibacterial chemotherapeutic agents, data on the use of PDF for other applications are extremely limited. The first application in biocatalysis was explored by Sonke and coworkers who described the enzyme-catalyzed stereoselective hydrolysis of racemic *N*-formylated molecules, the use of PDF in the stereoselective formylation of racemic amino acid derivatives, and the selective deprotection of *N*-formyl protected peptides [36]. The enzyme showed a high L-stereoselectivity for *N*-formyl amino acid derivatives, but the activities strongly depended on the substrate. Very low activity was found in case of bulky substrates, such as *N*-formyl-*tert*-leucine amide, as well as with α,α-disubstituted amino acid amides, i.e. *N*-formyl-(α-methyl)phenylglycine amide [36]. Moreover, in the case of deprotection of *N*-formyl peptides, major limitations were the low activity of *E. coli* PDF (EcPDF) towards formylated dipeptides with most amino acids in the first position [32] and the lack of an industrially applicable purification method.

From an industrial point of view, the application of Fe^{2+}-containing peptide deformylases in biocatalytic transformations is also limited by their intrinsic instability, which is mainly caused by the oxidation of the Fe^{2+} to Fe^{3+}. This can be efficiently prevented by the use of reducing agents (e.g. tris(2-carboxyethyl)phosphine (TCEP)) and hydrogen peroxide scavengers (catalase), or by substituting the Fe^{2+} metal cofactor with an ion (e.g. Ni^{2+}) that is less prone to oxidation [3;33]. Although the instability of the PDF can in principle be solved this way, the use of a scavenger and/or the replacement of the metal cofactor add to the cost of enzyme production and the complexity of the biocatalytic process. Therefore, the availability of PDF variants that are highly active and stable under aerobic conditions would be of great advantage to reduce costs and facilitate their industrial use. In any case, properties such as stability, kinetic parameters, and enantioselectivity have to be evaluated for such variants in order to decide if a new biological catalyst can be competitive with a chemocatalytic route or the use of an existing enzyme variant.

The question then is: how to find enzymes with ideal characteristics for a given biocatalytic process? The search for a new enzyme with specific properties is usually done either by the use of protein engineering to improve a known enzyme, or through the search for a better enzyme from the natural biodiversity. The choice for one or the other method is based on the specific characteristics that the biocatalyst should possess and the availability of a suitable starting enzyme. In fact, although natural diversity can provide a vast number of novel genes and activities relevant for industrial

applications [12], evolutionary selection in Nature never occurred to provide an optimal biocatalysts for an industrial application, making it unlikely that a natural enzyme is in any way optimal for a practical process. Therefore, the two aforementioned methods for obtaining improved enzyme variants are complementary, rather than alternatives. A new enzyme isolated from Nature will often need to be improved in the laboratory before it can be applied industrially.

Natural diversity can be searched in different ways: via the exploration of known cultures; through the use of enrichment cultures; by construction and screening of metagenomic libraries; or by database mining of genome sequences. The first three methods are time consuming and and costly. Furthermore, it is sometimes difficult to find a link between the enzymatic conversion of interest and bacterial growth. Since genome mining is a quick way to obtain a variety of cloned enzymes variants, we started with the selection and partial biochemical characterization of a smaller set of peptide deformylases encoded by genes from organisms available in culture collections. The respective PDF sequences can be recovered from genome sequence databases by homology searches. This approach is rapidly gaining potential in view of the vast amount of sequence data becoming available through next generation sequencing technologies. Accordingly, this Chapter describes the screening of genomic databases for peptide deformylase encoding genes and their phylogenetic analysis. Eight bacterial peptide deformylases were selected and their genes were cloned and expressed in *E. coli*, followed by assessing their properties with the aim to find a PDF with better activity, enantioselectivity and stability as compared to PDF from *E. coli* (EcPDF).

Another important issue addressed in this Chapter is the stability of PDF in the presence of cosolvents. In fact, the poor substrate solubility in water and the low activity of EcPDF in non-aqueous solvents are limiting its use in industrial processes. Since thermostability and solvent stability may be correlated, we selected the peptide deformylases from the termophile *Bacillus stearothermophilus* (BsPDF) for exploring its biocatalytic potential.

Materials and methods

Phylogenetic analysis of peptide deformylase genes

To obtain putative PDF sequences we performed BLAST searches using the *E. coli* PDF sequence as a query and the databases from the NCBI website (http://www.ncbi.nlm.nih.gov) and the Pedant database (http://pedant.gsf.de/).

A selection of 260 sequences resulting from BLAST hits (cutoff 20% identity) were aligned using ClustalW and the resulting phylogenetic tree was displayed using TreeView and edited.

PDF cloning and cultivation conditions

For all plasmid constructs described in this study, *E. coli* TOP10 was used as the host organism. It was routinely grown on LB medium supplemented with 100 mg/L of carbenicillin (Carb) at 20, 28 or 37°C. When protein expression was required from the *araBAD* promoter, L-arabinose was added at a concentration of 0.002, 0.02 or 0.2 wt % when the OD_{620} of the culture reached 0.5-0.8.

The peptide deformylase genes from *H. influenzae* (U32745)*, P. aeruginosa* (AE004441)*, H. pylori* (AE000591)*, E. coli* (AE00407)*, B. subtilis, Synechocystis sp.* (D90906)*, C. jejuni* (AL139074) and *B. stearothermophilus* (O31410) were amplified by PCR using specific primers and inserted in pDONR221 as per Gateway Cloning Technology protocols (Invitrogen, Breda, the Netherlands). The peptide deformylase genes were subsequently recombined to expression vector pBAD/*Myc*-His-DEST in a standard LR recombination reaction (Invitrogen) with a molar ratio destination to entry vector of 1:2.4 (150:300 ng). After transformation of *E. coli* TOP 10, recombinants were selected by plating on LB-plus-Carb (100 mg/L) medium, followed by overnight incubation at 28°C. Finally a correct clone for each gene, as established by plasmid DNA isolation and restriction analysis, was selected for further use. The expressed proteins do not contain a His-tag.

Protein determination and electrophoresis

Protein concentrations were determined by the Bradford method [6] with bovine serum albumin (Pierce, Rockford, Il) as the standard.

Protein electrophoresis was performed using sodium dodecyl sulfate-polyacrylamide (4 to 12%) Bis-Tris gels (NuPAGE, Invitrogen) under reducing conditions with MOPS-SDS running buffer (Invitrogen).

Determination of peptide deformylase activities

Deformylase activities of PDF prepations were routinely measured at 30°C in a total volume of 500 µL with *N*-formyl-Met-Ala-OH as the substrate. Amounts of 50 µL of different dilutions of PDF in a buffer containing 20 mM MOPS/NaOH (pH 7.7), 100 mM NaCl, 10 µg/mL bovine liver catalase, and 2 mM TCEP, were mixed with 450 µL of substrate solution containing 5.5 mM *N*-formyl-Met-Ala-OH in 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, and 0.1 mg/mL bovine liver catalase. The amount of PDF added was chosen such that in 15 min about 10% of the substrate was converted. After starting the reaction by addition of the PDF, samples of 100 µL were withdrawn from the reaction mixture with 5 min intervals and mixed with 100 µL of 1 M phosphate buffer (H_3PO_4-NaOH, pH 2.66) to stop the reaction. Substrate and product concentrations were measured using the HPLC as described below. To obtain an enzymatic blank, 500 µL of the same phosphate stopping buffer and 50 µL of diluted PDF were mixed prior to the addition of 450 µL of the substrate solution.

For determination of the substrate specificity of PDFs, incubations were done similarly. Samples of 50 µL of different dilutions of PDF (see above) were mixed with 450 µL of a 13 mM substrate solution in 110 mM MOPS/NaOH buffer (pH 7.2) containing 300 mM NaCl and 0.1 mg/mL of bovine liver catalase. After starting the reaction by adding PDF, 100 µL aliquots were withdrawn from the reaction mixture with 5 min time intervals, added to 400 µL of 50% ethanol in water, and analyzed by HPLC.

HPLC analyses

HPLC was carried out using a stainless-steel analytical column (250 mm length, 4.6 mm ID) packed with Inertsil ODS-3 material, 5 µm particle size, from Alltech Applied Science (Breda, The Netherlands). The flow rate was 1 mL/min. Detection was performed at 40°C at a wavelength of 210 nm. The injection volume was 5 µL.

For PDF deformylation activity measurements, For-Met-Ala-OH and H-Met-Ala-OH were separated using the following gradient of acetonitrile in 10 mM H_3PO_4: t=0-4 min, 0.1 % (v/v) acetonitrile isocratic; t=4-15 min, 0.1%-50% acetonitrile linear increase; t=15-15.1 min, 50%-0.1% linear decrease; t=15.1-20 min, 0.1% acetonitrile isocratic.

Quantification of formic acid was performed using pre-column derivatization of carbonic acids with 4-nitrophenylhydrazine (NPH) and 1-ethyl-3-[3-dimethylamino-propyl]carbodiimide (EDC) followed by HPLC. Separation was achieved on a reverse phase column with a gradient of acetonitrile and derivates were monitored at 400 nm [30].

For peptidase activity measurements, the production of Phe was analyzed using the following gradient of acetonitrile in 10 mM H_3PO_4: t=0-15 min, 0-30% acetonitrile linear gradient; t=15-15.1, 30-0% acetonitrile decrease; t=15.1-20, 0 % acetonitrile isocratic.

Temperature stability and optimum of PDF

The temperature stability of the enzymes was analyzed by preincubation of 0.05 mg/mL of cell-free extract containing an overexpressed PDF enzyme for 1 h at temperatures between 0 and 60°C. Subsequently, 50 µL samples of the pretreated enzyme solutions were used to determine the remaining activity according to the standard PDF assay at 30°C.

For the determination of the temperature optimum, the standard protocol for activity assays was applied at a temperature range from 20 to 60°C.

Influence of organic cosolvents on enzyme activity

The effect on enzyme stability of several water-miscible organic solvents [methanol, tetrahydrofuran (THF), dimethylformamide (DMF), tert-butanol, dimethyl sulfoxide (DMSO) and N-methyl-2-pyrrolidone (NMP)] (v/v) concentration was measured as follows. At 30°C, 50 µl samples of EcPDF and other PDF variants, pre-diluted in a solution containing 20 mM MOPS/NaOH (pH 7.7), 100 mM NaCl, 10 µg/mL bovine liver catalase, and 2 mM TCEP, was mixed with 450 µL of substrate solution composed of 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase, 5.5 mM of the substrate *N*-formyl-Met-Ala-OH and a final concentration of cosolvent of 10, 20 or 40%. Samples were periodically withdrawn from the incubation mixture to measure the amount of deformylated peptide formed. The residual specific activity is reported as a percentage of the specific activity of the enzyme that was found when no cosolvent was added to the reaction mixture.

Results and discussion

Phylogenetic analysis

The widespread distribution of phylogenetically related peptide deformylases in prokaryotic and eukaryotic organisms is evident from the detection of *E. coli* peptide deformylase (EcPDF) homologs in essentially all sequenced genomes. To identify bacterial peptide deformylases of varying origin that may possess enhanced stability and different substrate selectivity as compared to the EcPDF, we screened various databases for the presence of peptide deformylase-encoding sequences. In total, 260 annotated peptide deformylase sequences from different organisms were retrieved, using the sequence of EcPDF as a query. These sequences were subsequently aligned and a phylogenetic tree was generated using the program TreeView. This analysis clustered the sequences into 9 groups (Fig. 1).

The relatively long nodes as compared to the branches of the phylogenetic tree indicate low overall identity, even within a group. This is in agreement with the notion that only a few residues are crucial for proper deformylase activity allowing significant sequence variation. Despite the low sequence identity, the presence and topology of the amino acids involved in catalysis are well conserved, suggesting that the variants possess identical catalytic mechanisms and similar selectivities. This is also in agreement with the fact that the amino acid sequences of the peptide deformylases from *E. coli* PDF and *S. aureus* PDF are only 23% identical, but an examination of their structures in the vicinity of the metal ion in the active site revealed no significant differences [2].

PDFs are often divided in two different groups: type I and type II enzymes [13;16;18]. Type I enzymes would generally occur in Gram-negative bacteria whereas group II enzymes would occur in Gram-positive bacteria (red line in Fig. 1, [13] However, there is no strict correlation between the type of PDF that is present in a certain organism and its classification (i.e by Gram reactivity) and some bacteria have both a type I and a type II enzyme. An analysis of genome databases shows that bacteria can contain one (e.g *E. coli*), two (*B. stearotermophilus, E. faecium, E. faecalis*), or up to four (*S. coelicolor*) PDF gene sequences (Fig. 1). In general, gram-negative organisms have a single PDF gene, while gram-positive bacteria, including *S. aureus*, typically carry two, although one of them often encodes an inactive protein [16].

Specific structural features distinguish the type I and type II bacterial PDFs. Type II enzymes are characterised by the presence of two sequence insertions in the N-terminal part of the enzyme and by the presence at the *C*-terminus of a long loop which ends with a small β-strand that folds back in the direction of the second β-sheet of the enzyme. In type I enzymes, the N-terminal part is shorter and an α-helix is present at the *C*-terminus [16]. Despite the differences in secondary structure between enzymes of the two groups, the conformation of the active site and the catalytic properties are very similar.

The enzymes of groups I and II cluster in phylogenetically different subgroups. As mentioned, type I PDFs are generally present in Gram-negative bacteria (*proteobacteria*, *cyanobacteria*, *Chlamydiae*, *Spirochaetales*) and in eukaryotes (see Fig. 1), whereas type II enzymes appear in two clusters, one formed by PDFs from Firmicutes and the other by PDFs from Actinobacteria. An exception to this classification is for example the PDF from *M. tuberculosis* (MtPDF). MtPDF has features similar to type I PDFs, with an extended *C*-terminal portion but the enzyme also has several insertions typical of type II enzymes [31].

Type I PDFs can be divided in two subclasses: PDFIA and PDFIB. The first subclass corresponds to animal- and plant-mitochondrial enzymes, while the PDF1B enzymes occur in gram-negative bacteria, some gram-positive bacteria and plants [35]. PDFs from eukaryotes, including plant mitochondrial and chloroplast PDFs, as well as PDFs from protozoa, are classified as type I PDFs since they do not have the two typical insertions of type II enzymes and occur in a separate subgroup. While PDFIB is similar in its 3D structure and biochemical properties to type II bacterial PDFs, PDFIA has as main difference the nature of the metal cofactor. *Arabidopsis thaliana* PDF (AtPDF) uses Zn^{2+} as cofactor, probably reflecting the need to be active in the highly oxidative environment of plant mitochondria [34]. Human mitochondrial PDF (HsPDF) is a dimer and has a tetrahedrally coordinated Co^{2+} [11]. It possesses about 30% sequence identity with bacterial and plants PDFs and the degree of similarity is about 50%, which is reflected in a conserved α/β-fold. The overall fold resembles the one of EcPDF (a type IB PDF), although it possesses a more accessible entrance to the active site due to a different set of residues that in EcPDF occlude the entrance site [11]. The in vivo role of the type I eukaryotic PDFs is not completely clear although in recent years an enzymatic machinery for the formylation and deformylation of proteins was found both in mammals and in plants [5;10;34]. Animal PDFs are similar to the enzymes found in

Actinobacteria, even though mitochondrial genes are believed to originate from Rickettsiae [19]. Therefore, a possible hypothesis of their origin is horizontal gene transfer from an actinobacterium to an ancestor eukaryotic cell [35].

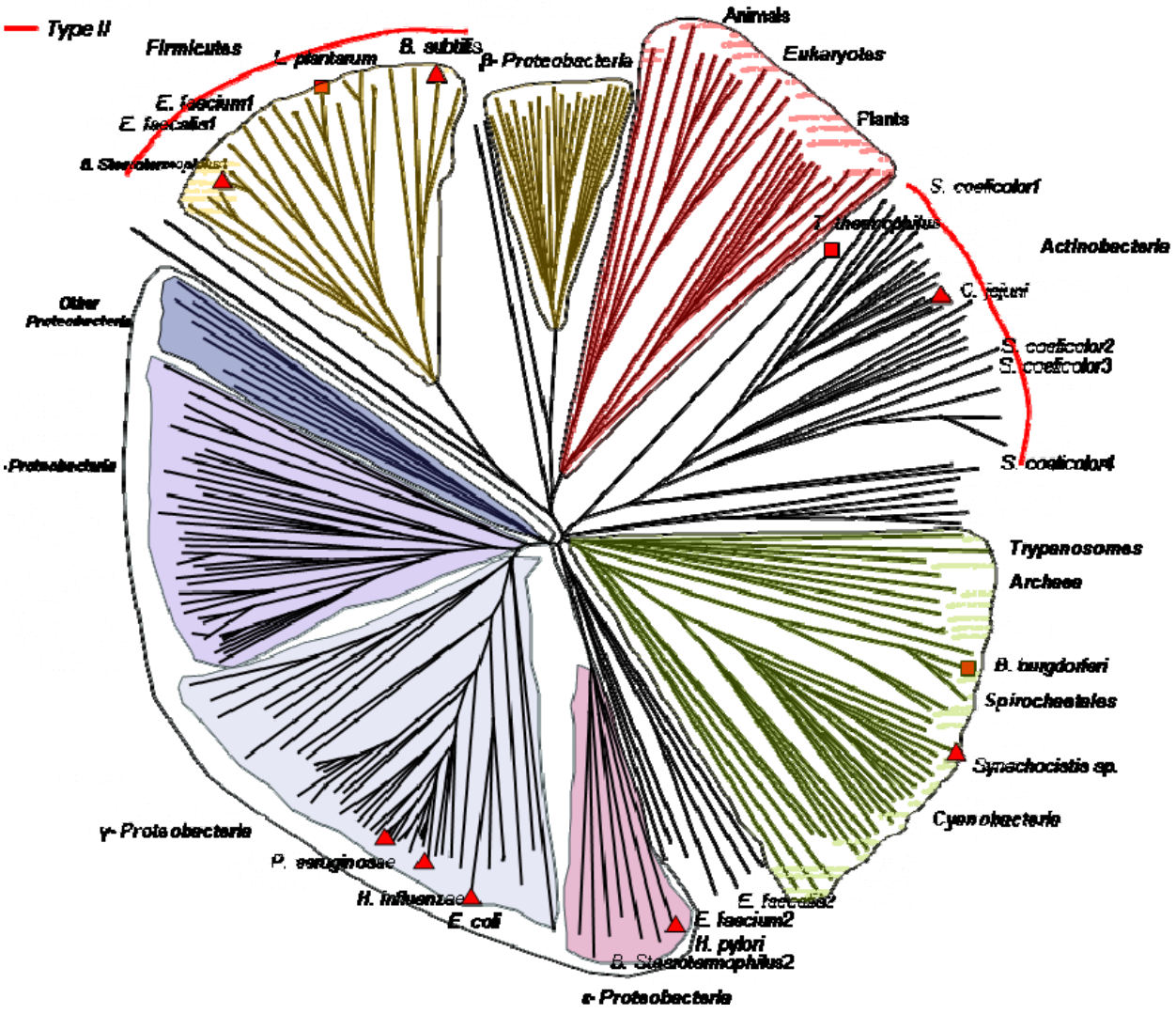


Figure 1. Phylogenetic tree of peptide deformylases. Symbols: squares, PDFs from *T. thermophilus, B. burgdorferi* and *L. plantarum*; triangles, PDFs from *H. influenzae, P. aeruginosa, H. pylori, E. coli, B. stearothermophilus, Synechocystis sp., B. subtilis,* and *C. jejuni.*

Eukaryotic PDFs are also characterized by an extension at the *N*-terminus and a longer CD-loop between strands β_2 and β_3. This loop can extent close to the active site and may have a role in changing its conformation, as reported for AtPDF and the *Plasmodium falciparum* enzyme (PfPDF) [9;28]. Studies conducted on animal PDFs showed a marked difference compared to all other PDFs by the presence of conserved

substitutions in the vicinity of the active site. In fact, the first Gly and the second Ala in conserved motif 1 (sequence GIGLAATQV in EcPDF) and Leu in motif 2 (EGCLS in EcPDF) are replaced by other residues and this is responsible for the low activity of the mitochondrial PDFs [35].

A

```
L. brevis              1   ------------MFLMKDIVRDGDPVLRQEAADVTFPLSEEDQQLAKDLMEYLVVSQDPEQCKKYGLRAGVGLAAPQVGV  68
B. cereus              1   ------------MLTMKDVIREGNPILRNVAEEVSLPASEEDTTTLKEMIEFVINSQDPEMAEKYSLRPGIGLAAPQIGV  68
L. plantarum           1   ------------MIKMRDIIREGNHTLRAEAKQVKFPLSEADQKLANDMMEYLENSQDPELAKKYGLRAGVGLAAPQVDV  68
B. stearothermophilus  1   ------------MITMKDIIKEGHPTLRKVAEPVPLPPSEEDKRILQSLLDYVKMSQDPELAAKYGLRPGIGLAAPQINV  68
E. faecalis            1   ------------MITMKDIIREGNPTLRAVAEEVPVPITEEDRQLGEDMLTFLKNSQDPVKAEELQLRGGVGLAAPQLDI  68
S. pneumoniae          1   MSAIERITKAAHLIDMNDIIREGNPTLRTVAEEVTFPLSDQEIILGEKMMQFLKHSQDPVMAEKMGLRGGVGLAAPQLDI  80
S. hominis             1   ------------MLTMKDIIRDGHPTLREKAKDVNLPLSEEDKNTLRDMREFLINSQDDEIAKKYGLRSGVGLAAPQINI  68
S. aureus              1   ------------MLTMKDIIRDGHPTLRQKAAELELPLTKEEKETLIAMREFLVNSQDEEIAKRYGLRSGVGLAAPQINI  68
E. coli                1   ------------MSVLQVLHIPDERLRKVAKPVE-EVNAEIQRIVDDMFETMYAE------------DGIGLAATQVDI  54
```

B

	motif 1 ######### GXGXXAXQX		motif 2 ##### EGCLS		motif 3 ####### QHEXDHL	
E. coli	GIGLAATQVDIHQRIIV	59	IEEGCLSIP-EQ	96	IQHEMDHLVG	139
H. influenza	GIGLAAPQVDILQRIIT	60	IEEGCLSIP-GF	97	IQHEIDHLNG	140
P. aeruginosa	GIGLAATQVNVHKRIVV	60	YQEGCLSVP-GF	98	IQHECDHLNG	141
H. pylori	GIGLAAIQVGLPLRMLI	59	YKEGCLSVP-GF	101	IQHEIDHLNG	144
B. subtilis DEF1	GVGLAAPQIGILKRAAV	60	GIEGCLSFP-NV	96	VQHEMDHLDG	139
B. subtilis DEF2	GVGLAAPQINIKKRMIA	74	SGEGCLSVDEAI	117	FQHEIDHLNG	160
B. cereus	GIGLAAPQIGVSKKMIA	75	GGEGCLSVDREV	118	FQHEIDHLNG	161
B. stearothermoph	GIGLAAPQINVSKRMIA	74	TGEGCLSVDRDV	117	FQHEIDHLNG	160
Synechocystis	GIGLAAPQVGINKQLLV	73	VEEGCLSVP-NV	113	IQHEMDHLNG	156
Thermotoga	GVGLAAPQVGISQRFFV	70	AEEGCLSFP-EI	105	FQHEFDHLNG	148
Plasmodium	GIGLSAPQVNISKRIIV	59	LIEGCLSFP-GI	100	FQHEFDHLNG	143
Arabidopsis PDF1A	GVGLAAPQIGVPLRIIV	140	FFEGCLSVD-GF	194	LQHECDHLDG	237
L. brevis	GVGLAAPQVDVSKQMSA	74	EGEGCLSVDREV	119	CQHEIDHLNG	162
Borrelia	GVGLAAPQVGLDLALFV	64	YKEGCLSIP-GV	101	IQHEMDHLNG	144
S. pneumoniae	GVGLAAPQLDISKRIIA	86	EGEGCLSVDRNV	137	VQHEIDHING	180
C. jejeuni	GVGLAAIQVDIPLRVLL	55	CTEGCLSVP-DF	102	IQHENDHLNG	145
Leptospira	GVGLAAPQIGILKQIVV	62	FWEGCLSVP-GM	107	YQHECDHLQG	150
M. tuberculosis	GVGLAANQIGCSLRLFV	65	DDEGCLSVP-GE	113	LQHETGHLDG	155
Trypan. brucei TbPDF1	CISFSAPKGHWDAAIVL	98	MWENCISCGACT	143	LMHELDHLSG	185
Trypan. brucei TbPDF2	YPSLCAPKIGWNVQMMF	127	AWEPCASCCFLL	86	ALHELDHLNG	229
Homo sapiens mitoch	CVGLSAPQLGVPRQVLA	124	FPEGCESVA-GF	179	IQHEMDHLQG	221
Mus musculus mitoch	CVGLSAPQLGVPLQVLA	102	FPEGCESVAGFL	166	IQHEMDHLQG	109

Figure 2: Comparison of PDF sequences. A, Sequence alignment of N-terminal region of selected PDFs. The insertion present in the AB loop of type II PDFs is indicated with a box. B, Conserved sequence motifs in PDFs. See text for details.

In the last decade a third class of PDFs was identified in trypanosomes and *Archaea*. These PDFs are placed in a cluster from which originate also PDFs of *Cyanobacteria*, *Spirochaetales* and some eukaryotes, including protists and amoeboids (light green in Fig. 1). Studies on these PDF sequences revealed that they have mutations in another highly conserved motif, motif 3 (e.g. lacking the Q that is present in *E. coli* QH*E*MDHL), suggesting that these PDFs may have different catalytic properties (Fig. 2). Moreover, the phylogenetic tree highlights two groups which are more distant at the sequence level. The organisms with these PDFs are Gram-positive bacteria belonging to the Firmicutes and Gram-negative β-proteobacteria (yellow in Fig. 1).

The iron-free PDFs from bacteria growing under strict iron limitation (*L. plantarum*, *Borrelia*) do not appear in a separate cluster, but are similar to PDFs of phylogenetically related organisms (*Fermicutes* and *Spirochaetales*, respectively) (Fig. 1).

Table 1. Activity of various PDF enzymes .

Microorganism	Abbreviation	Sp. Act. (U/mg)[a]	Sp. Act. purified enzyme (U/mg)
H. influenzae	HiPDF	120	$2.1 * 10^3$
P. aeruginosa	PaPDF	155	$4.3* 10^3$
H. pylori	HpPDF	60	$1.3 * 10^3$
E. coli	EcPDF	370	$4 * 10^3$
B. stearothermophilus	BstPDF	20	130
Synechocystis sp.	SyPDF	70	650
B. subtilis	BsPDF	25	270
C. jejuni	CjPDF	20	200

[a] Measured with For-Met-Ala-OH as substrate.

Overexpression and purification of peptide deformylases

Eight peptide deformylases-encoding genes from different phylogenetic groups were selected for cloning and overexpression in *E. coli* (Table 1). The genes were amplified from chromosomal DNA and cloned into a pBAD-derived expression vector as described under Materials and Methods. Expression levels were tested by activity measurement and SDS-polyacrylamide gel electrophoresis (Fig. 4-6). From the

sequence similarity studies, *H. influenzae* (HiPDF), *P. aeruginosa* (PaPDF) and *H. pylori* (HpPDF) deformylase genes showed the highest sequence similarity with the EcPDF *def* gene (triangles in Fig. 1) and this was clearly reflected by the best heterologous expression in *E. coli*. The BsPDF, SyPDF, PaPDF, HiPDF and HpPDF were present at a level of 10-30% of the total soluble protein (Fig. 3) while expression of BstPDF and CjPDF was rather low.

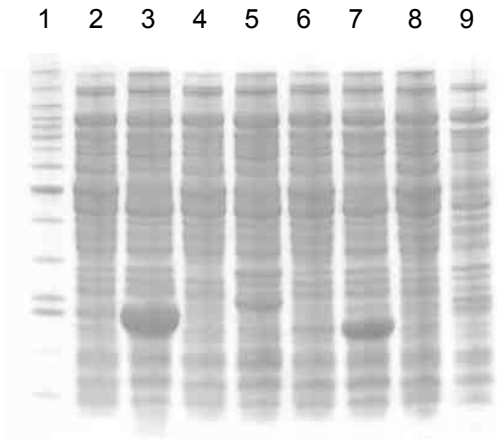


Figure 3. SDS-PAGE of CFE from induced and non-induced cells.
Lane 1, reference protein ladder;
Lane 2, not induced EcPDF;
Lane 3, induced EcPDF
Lane 4, not induced CjPDF;
Lane 5, induced CjPDF,
Lane 6, not induced BsPDF,
Lane 7, induced BsPDF,
Lane 8, not induced BstPDF,
Lane 9, induced BstPDF.

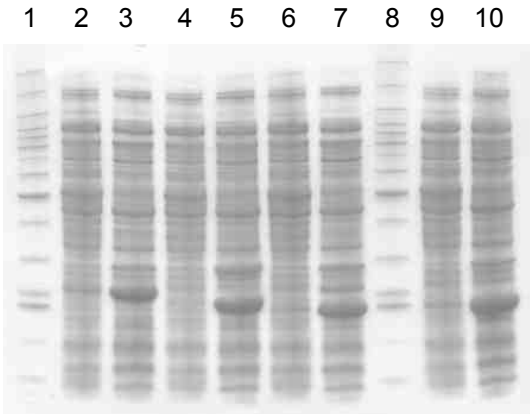


Figure 4. SDS-PAGE of CFE from induced and non-induced cells
Lane 1, reference protein ladder;
Lane 2, not induced SyPDF;
Lane 3, induced SyPDF,
Lane 4, not induced PaPDF;
Lane 5, induced PaPDF,
Lane 6, not induced HpPDF,
Lane 7, induced HpPDF,
Lane 8, reference protein ladder,
Lane 9, not induced HiPDF,
Lane 10, induced HiPDF.

Analysis of the CFE prepared from an expression culture of the BsPDF in *E. coli* TOP10 showed that good overexpression occurred, even though the specific activity was rather low (Fig. 3-5).

To isolate and characterize the overexpressed enzymes, affinity column chromatography as described by Sonke and coworkers was used [36]. The use of this chromatographic method is based on the Met-Lys-Sepharose affinity matrix and gives pure protein in a single step. Other attractive features of this method are the fact that native proteins can be isolated (affinity tags are not necessary) and that it can be universally applied. The PDF enzyme binds to the Met-Lys-Sepharose affinity matrix in the presence of fluoride ions, which dramatically increase the affinity of the enzyme for small peptides, most likely because fluoride mimics formate. Switching to a chloride-containing buffer decreases the affinity of PDFs for the ligand causing the elution of the desired enzyme. In addition, this affinity method allows the purification only of the active form of the peptide deformylase.

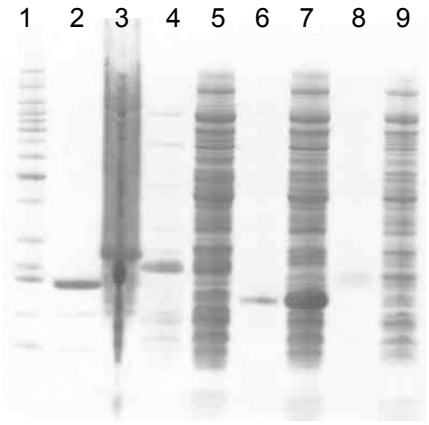


Figure 5. SDS-PAGE of CFE and purified PDFs obtained with Met-Lys-Sepharose affinity column chromatography.
Lane 1, reference protein ladder;
Lane 2, purified. EcPDF;
Lane 3, EcPDF CFE
Lane 4, purified CjPDF;
Lane 5, CjPDF CFE,
Lane 6, purified BsPDF,
Lane 7, BsPDF CFE,
Lane 8, purified BstPDF,
Lane 9, BstPDF CFE.

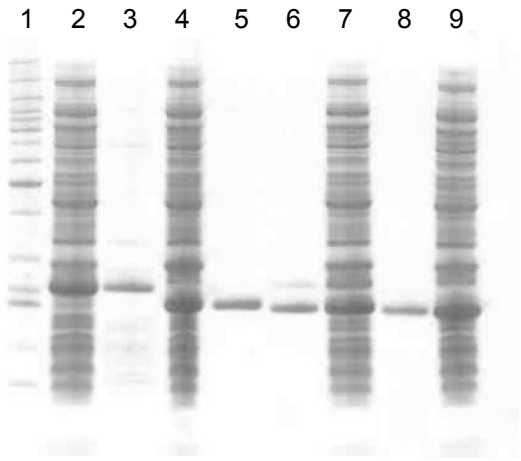


Figure 6. SDS-PAGE of CFE and purified PDFs obtained with Met-Lys-Sepharose affinity column chromatography.
Lane 1, reference protein ladder;
Lane 2, SyPDF CFE;
Lane 3, purified SyPDF
Lane 4, PaPDF CFE;
Lane 5, purified PaPDF,
Lane 6, purified HpPDF,
Lane 7, HpPDF CFE,
Lane 8, purified HiPDF,
Lane 9, HiPDF CFE.

The affinity purification method was successfully applied to all PDFs that were overexpressed, which resulted in an increase of specific activities by a factor 10 to 30 (see Table 1). Thus, eight different PDFs were isolated in a single purification step, and all were catalytically active.

Effect of temperature on enzyme stability

The thermal stability of the purified PDFs was determined by measuring the remaining activity after preincubation of the purified enzyme at a concentration of 3 µg/mL for 60 min at temperatures between 20 and 80 °C. The activity of all PDFs decreased dramatically with the increase of temperature, with a loss of 50% after incubation at temperatures ranging between 30 and 50°C. These unexpected results might be explained by the fact that PDFs were diluted to concentrations <50 µg/mL to mimic process conditions. It is known that peptide deformylase is highly unstable at low concentrations [15]. This explanation was confirmed by repeating the thermostability test for EcPDF at higher concentrations, i.e. 500 µg/mL instead of 3 µg/mL (Fig. 7). The undiluted sample of EcPDF appeared to be stable at 30°C whereas a decrease of 50% was observed when the diluted enzyme was preincubated at30°C (Figure 7).

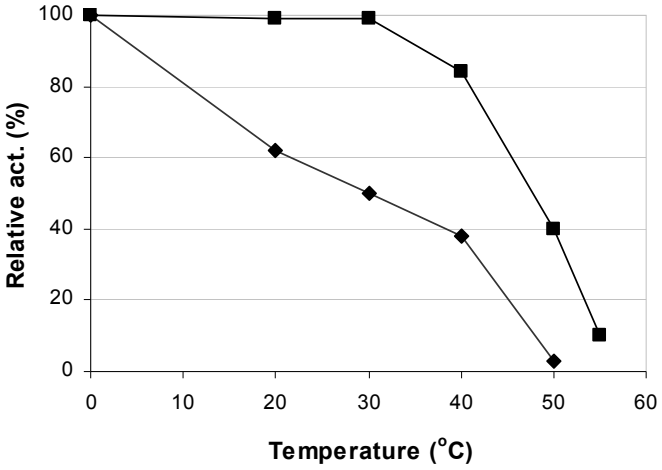


Figure 7. EcPDF thermostability; relative activity after 60 min pre-incubation at increasing temperatures. Effect of PDF concentration. Symbols:

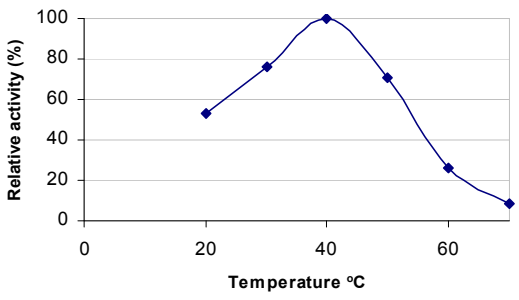
♦ EcPDF (3 µg/mL)

■ EcPDF (500 µg/mL)

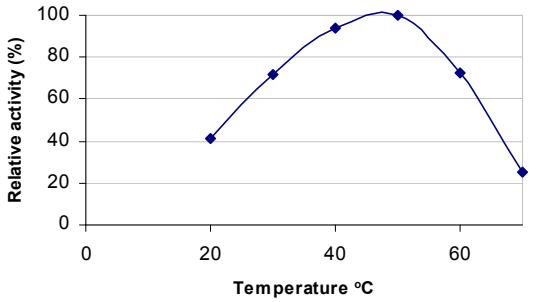
Temperature optimum

A higher temperature will increase the reaction rate but may also enhance enzyme inactivation. Therefore a proper evaluation of enzyme performance under process conditions requires measurement of the temperature optimum and stability. Activity measurements were performed with all purified peptide deformylases at temperatures between 20 and 70°C. BstPDF and HiPDF showed a higher temperature optimum (45 to 50°C) while CjPDF, HpPDF and SyPDF showed highest activities at around 25°C (Figure 8).

A



B



\

C

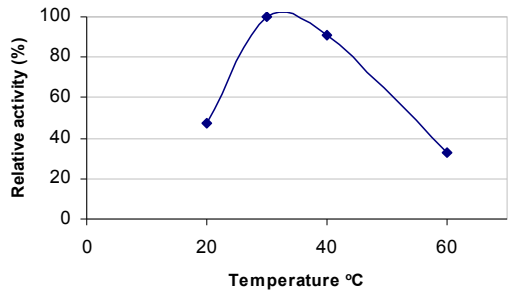


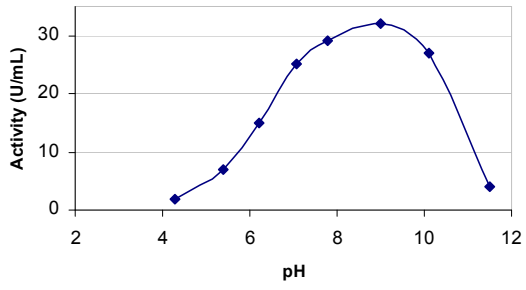
Figure 8. Temperature optimum of BstPDF (A), HiPDF (B) and EcPDF (C)..

Effect of pH on enzyme activity

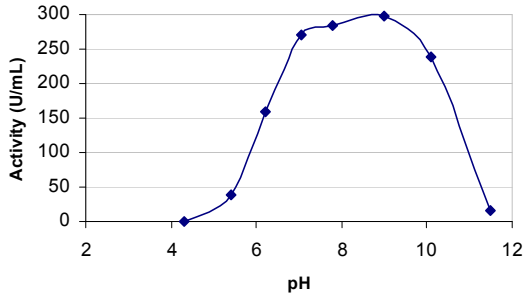
Enzyme activities were determined between pH 4.0 and 11.5 with a universal Britton-Robinson buffer, because the use of different buffers over a large pH range led to buffer-dependent effects. The purified enzymes were active towards For-Met-Ala-OH in the whole range of tested pH values (Fig. 9). The pH optimum varied among enzymes from different species: BstPDF and BsPDF showed maximum activity at pH 9 while EcPDF and SyPDF performed better at pH around 7.

Overall, the PDFs showed near maximum activity over a broad pH range, from about pH 7 to 10, probably due to the absence of ionisable groups in the oxyanion hole [4]. At pH values of about 5 and 11, however,, the peptide deformylases had only about 50% of their maximum activity.

A



B



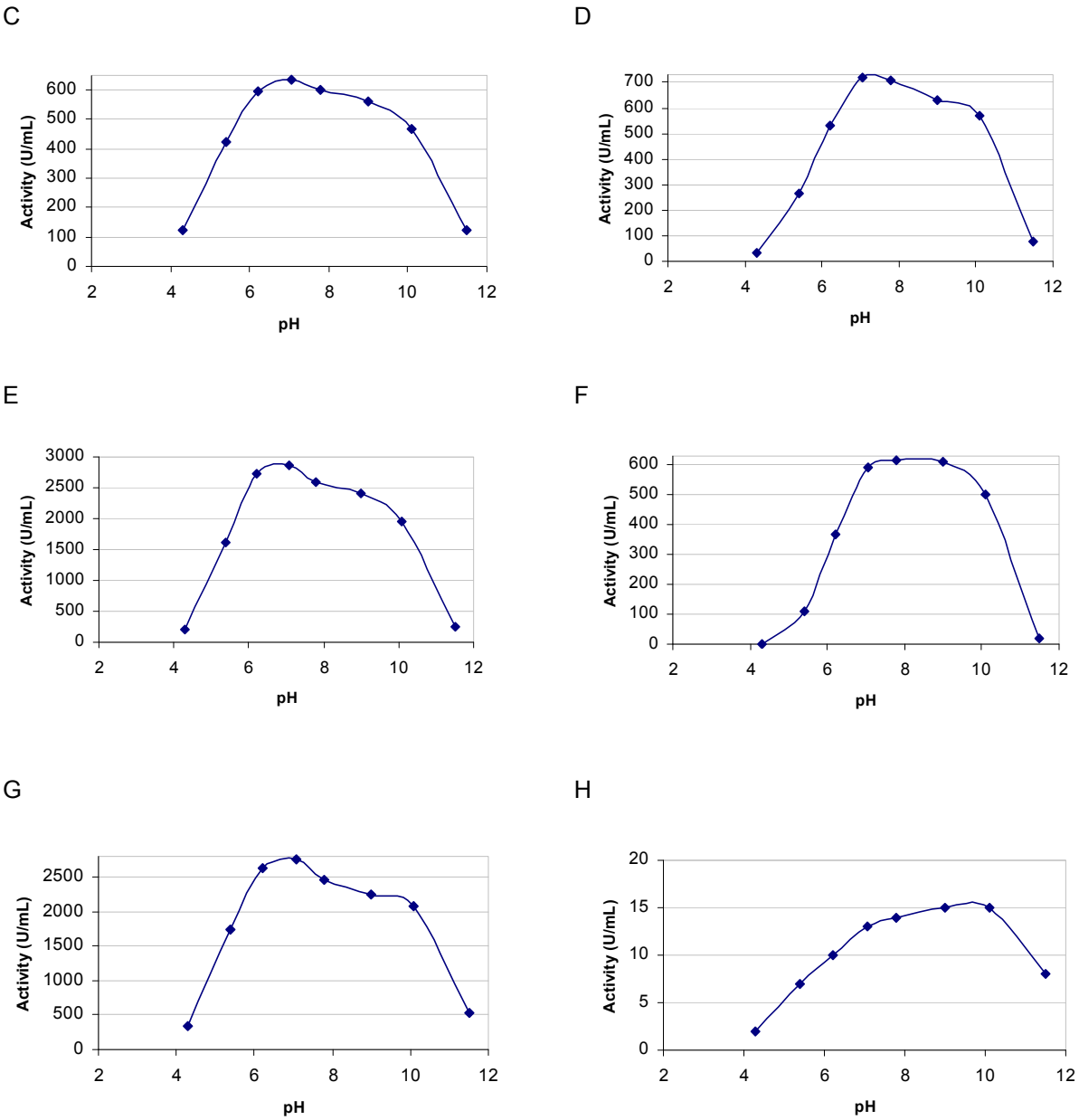


Figure 9. Effect of pH on activity of different PDFs. Enzymes: BsPDF (A), CjPDF (B), HiPDF (C), HpPDF (D), PaPDF (E), SyPDF (F), EcPDF (G), BstPDF (H). See text for details.

Effect of organic cosolvents on enzyme stability

As described in Chapter 2, the insolubility of peptide intermediates and protected peptides in water is a major problem for peptide synthesis. Polar organic solvents such as dimethyl sulfoxide (DMSO), dimethylformamide (DMF), methanol (MeOH), N-methylpyrrolidone (NMP), tetrahydrofuran (THF) and 2-methyl-2-propanol are often used as cosolvents to increase peptide solubility [27]. Unfortunately, previous data showed that EcPDF was quite unstable in the presence of organic cosolvents (Results and Discussion, Chapter 2). We tested BstPDF in the presence of different

concentrations of cosolvents since it has been demonstrated that a clear correlation exists between thermostable enzymes as derived from thermopiles and resistance to denaturation by the presence of cosolvents [29] [8]. In addition, thermophilic enzymes have longer storage stability and possess the ability to perform reactions at elevated temperatures. They also may be more stable under process conditions [1].

Unfortunately, the effect of MeOH and DMSO on BstPDF stability was similar to what was observed for EcPDF, with only about 50% activity remaining even when the cosolvent concentration was only 10% (Fig. 10).

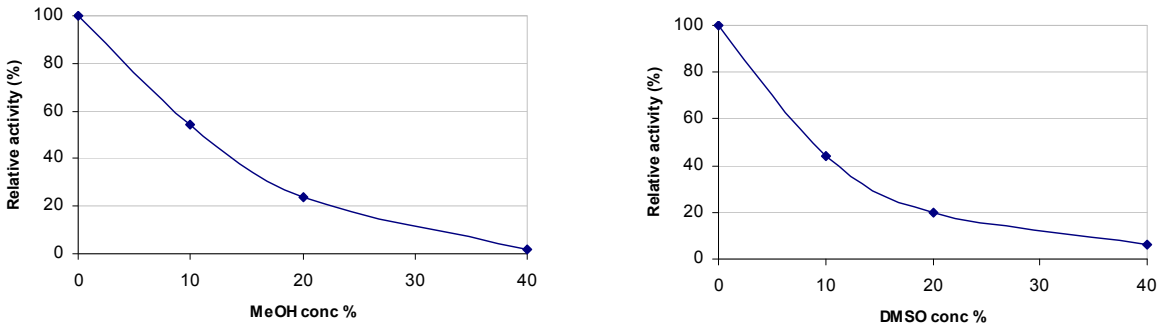


Figure 10. Effect of cosolvents on the activity of BstPDF.

Substrate specificity of peptide deformylases

The substrate specificity of the purified peptide deformylases was tested towards a range of racemic *N*-formylated amines and amino acid derivatives. All the reactions were performed with the standard assay at pH 7.2 and conversions were followed by measuring formic acid production as described in Materials and Methods. The advantage of this method is that it can be generally applied for every substrate of interest. Because of the low solubility of the substrates used, substrate stock solutions were prepared in dimethyl sulfoxide (DMSO) before addition to the reaction mixture. This cosolvent was chosen despite the stability issues mentioned in the previous paragraph, but the final concentration in the assays was kept below 10%.

The results in Table 2 show that the activity of the PDFs strongly varies for the different substrates. First it appears that with all enzymes the most reactive substrate

was the dipeptide *N*-formyl-Met-Ala-OH having a methionine in position P1. Furthermore, *N*-acetyl-phenylalanine nitrile was not hydrolyzed by any of the PDFs whereas *N*-formyl-phenylalanine nitrile was readily converted, confirming the previously reported observation that the enzyme is active only on *N*-formylated derivatives [25;36]. Inspection of the structure of EcPDF (PDB 1BS6) shows that the additional methyl group would overlap with the side chain of Leu-91 and Leu-46 and with the carbonyl oxygen of Gly-45 [7]. Two other substrates for which none of the PDFs tested displayed activity are *N*-formyl-α-methyl-benzyl amine and *N*-formyl-sec-butyl amine. This is probably due to the absence of hydrogen bond donors or acceptors in these molecules because a hydrogen bond between the substrate's formulated nitrogen and the carbonyl oxygen of Gly-45, as well as a hydrogen bond to O1 of Gly-89 are important for activity and can not be formed (Fig. 11B).

There are also differences between the activity for *N*-formyl-phenylalanine, *N*-formyl-phenylalanine-NH_2 and for *N*-formyl-phenylalanine nitrile. Also in this case, the lower activity towards *N*-formyl-phenylalanine could be caused by the inability of hydrogen bond formation between the substrate and Gly-89. Becker and coworkers proposed that this hydrogen bond, donated by the amide nitrogen of the second amino acid of the substrate (P2), is communicated to the metal ligand Cys-90, thereby reducing the energy of the transition state and providing a sort of substrate-assisted catalysis (Fig. 11B) [4].

Comparing the hydrolysis of the two N-formyl α-aminonitriles confirmed the importance of the side chain of the amino acid which has to fit into the hydrophobic methionine-binding pocket. The higher activity observed with *N*-formyl-phenylalanine nitrile is in line with previous observations, reporting a k_{cat}/K_m value that is 10-fold higher with *N*-formyl-Phe-$X_{aa}(NH_2)$ compared to *N*-formyl-Leu-$X_{aa}(NH_2)$ [32].

These experiments were performed with the aim to explore the substrate specificity of PDFs belonging to different phylogenetic classes and harvested from various microbial sources. The results showed that the substrate specificities of the tested PDFs are very similar, even though they originate from very different organisms and have significant sequence divergence.

Table 2. Substrate specificity of PDFs[a] from various sources expressed in *E. coli*.

Organism	N-formyl-Phe	N-formyl-Phe nitrile	N-formyl-Phe amide	N-formyl-Phe isopropylester	N-acetyl-Phe nitrile	N-formyl-Leu nitrile	N-formyl-α-methyl benzylamine	N-formyl-sec butylamine	N-formyl-Met-Ala
CiPDF	+	++	++	++	-	+	-	-	++++
EcPDF	+	++	+++	+++	-	++	-	-	++++
HiPDF	++	++	+++	nd[b]	-	+	-	-	++++
HpPDF	+	++	+++	nd	-	++	-	-	++++
PaPDF	+	+++	+++	nd	-	++	-	-	++++
SyPDF	+	+++	+++	nd	-	+	-	-	++++

[a] Activities were scored as follows: -, ≤ 10 U/g; +, 10-100 U/g; ++, 100-1000 U/g; +++, 1000-6000 U/g; ++++, ≥ 200 U/mg.

[b] nd, not determined.

Considering the importance of *N*-formyl methionine deformylation during ribosomal protein synthesis, it is not surprising that peptide deformylase optimally act on peptides having the sequence *N*-formyl-Met-X [17]. Structurally, the high specificity of the enzyme is due to the deep burial of the P1' side chain residue (methionine) in a hydrophobic cleft situated close to the active site (Fig. 11A). This interaction contributes to the affinity of PDF for its substrate and facilitates the initial nucleophilic attack on the formyl group by a water molecule that is bound to the metal ion in the PDF active site. Although other amino acid side-chains can enter the hydrophobic pocket, such substrates display a relatively low affinity (high K_m values) and the hydrolysis rates (k_{cat} values) can be very low. This is likely due to non-optimal positioning of the formyl moiety as a result of the forced insertion of a P1' side chain in the hydrophobic pocket[17].

Knowledge about substrate recognition and selectivity of PDFs is still limited. Extensive studies were only reported by Meinnel and coworkers, who showed the effect of the first and second amino acid residue in formylated dipeptide substrates on the activity of EcPDF [32]. Their results indicated that the presence of the hydrophobic pocket allows efficient deformylation of methionine only. The effect on the activity with amino acid derivatives has, to our knowledge, not been explored earlier, most studies focusing on inhibition of the enzyme by various complex molecules.

It is commonly assumed that enzymes evolve towards acquiring a higher catalytic efficiency to better accomplish their biological function. The substrate turnover rate and specificity of enzymes control the flux through interconnected metabolic pathways. Whether or not a stringent selectivity is required depends on the physiological function

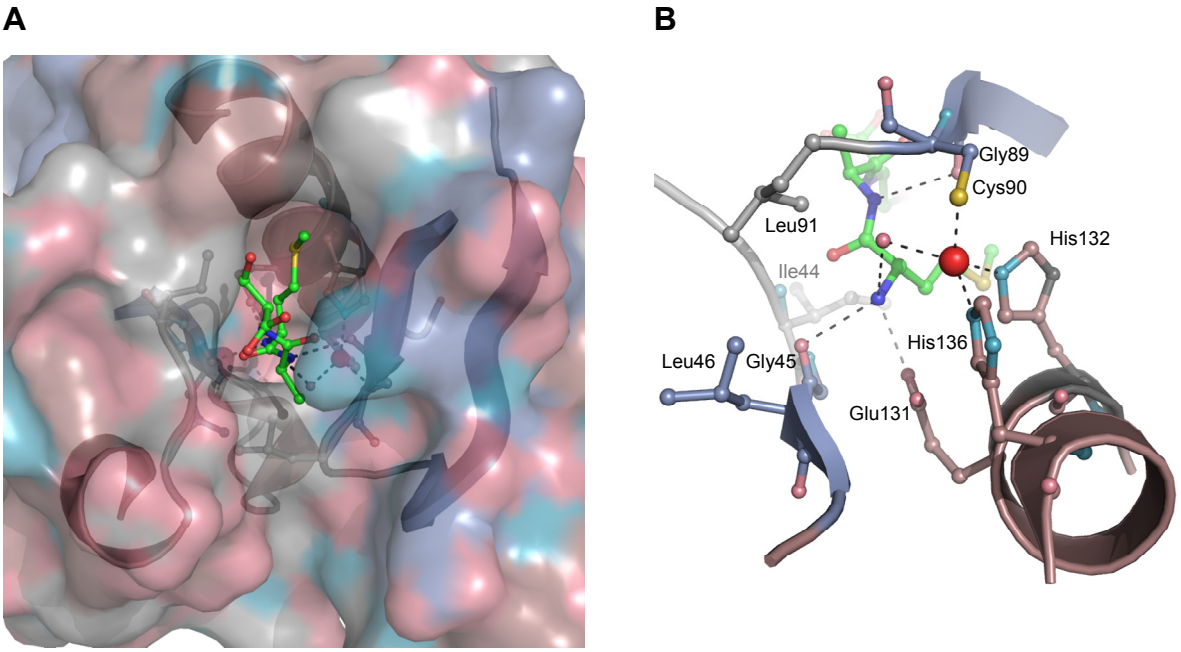


Figure 11. Active site pocket of EcPDF. Figures were generated using pdb structure 1BS6. Left: structure of the active site pocket shown in surface display, H-Met-Ala-Ser-OH in ball and sticks (green). Right: close up of the active site with H-Met-Ala-Ser-OH and Ni^{2+} binding interactions. Picture made with Pymol software.

of an enzyme [22]. For example, enzymes involved in catabolism may have a broad substrate range, while enzymes playing a role in central metabolism and anabolism must be specific in order to discriminate among metabolic intermediates having only small structural differences. Consequently, peptide deformylase and methionine aminopeptidase have evolved towards very narrow substrate specificity and a high

catalytic activity for their natural substrates. The substrate recognition sites of the different PDFs are structurally conserved and restricted, thus providing a similar stringent substrate specificity to enzyme variants even when these have a low sequence similarity and occur in phylogenetically very different organisms [21;37]. In view of these considerations and the results obtained, the chances to find natural peptide deformylase variants with broader substrate specificity can be regarded as very low.

Conclusions

A potential way to obtain peptide deformylases with higher stabilities and better catalytic properties than that of EcPDF is through mining of sequenced genomes. In this chapter we report the cloning and overexpression in *E. coli* of eight peptide deformylases from phylogenetically different organisms. The enzymes could be expressed in *E. coli* and were easily purified using the Met-Lys-Sepharose affinity chromatography method developed for the *E. coli* enzyme. The characterization of these enzymes revealed different temperature optima, with BstPDF and HiPDF having an optimum temperature at 40-50 °C and somewhat higher temperature stability. A major factor influencing the temperature stability was the concentration of the purified enzyme, which is in line with early studies of Groche and colleagues who showed that when stored at high concentrations EcPDF can remain active for several weeks [15]. The substrate specificity of the recombinant PDFs was also tested, but despite the choice of enzymes from very different organisms, no major differences in substrate range were discovered. In view of these results and the biological function of PDFs the chance to find in nature PDFs with a broader substrate range than that of the *E. coli* enzyme is regarded as small. An alternative method for obtaining improved peptide deformylases is to redesign a specific PDF, which is explored in the following Chapters.

References

1. **Cabral J.S.** (Ed) 2000. Applied biocatalysis. European Federation of Biotechnology. Working Party on Applied Biocatalysis. Harwood Academic Publishers.
2. **Baldwin, E. T., M. S. Harris, A. W. Yem, C. L. Wolfe, A. F. Vosters, K. A. Curry, R. W. Murray, J. H. Bock, V. P. Marshall, J. I. Cialdella, M. H. Merchant, G. Choi, and M. R. Deibel, Jr.** 2002. Crystal structure of type II peptide deformylase from *Staphylococcus aureus*. J. Biol. Chem. **277**:31163-31171.

3. **Becker, A., I. Schlichting, W. Kabsch, S. Schultz, and A. F. Wagner**. 1998. Structure of peptide deformylase and identification of the substrate binding site. J. Biol. Chem. **273**:11413-11416.

4. **Becker, A., I. Schlichting, W. Kabsch, D. Groche, S. Schultz, and A. F. V. Wagner**. 1998. Iron center, substrate recognition and mechanism of peptide deformylase. Nat. Struct. Mol. Biol. **5**:1053-1058.

5. **Bracchi-Ricard, V., K. T. Nguyen, Y. Zhou, P. T. R. Rajagopalan, D. Chakrabarti, and D. Pei**. 2001. Characterization of an eukaryotic peptide deformylase from *Plasmodium falciparum*. Arch. Biochem. Biophys. **396**:162-170.

6. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

7. **Chan, M. K., W. Gong, P. T. R. Rajagopalan, B. Hao, C. M. Tsai, and D. Pei**. 1997. Crystal structure of the *Escherichia coli* peptide deformylase. Biochemistry **36**:13904-13909.

8. **Cowan, D. A.** 1997. Thermophilic proteins: stability and function in aqueous and organic solvents. Comp. Biochem. Physiol. A **118**:429-438.

9. **Dirk, L. M. A., J. J. Schmidt, Y. Cai, J. C. Barnes, K. M. Hanger, N. R. Nayak, M. A. Williams, R. B. Grossman, R. L. Houtz, and D. W. Rodgers**. 2008. Insights into the substrate specificity of plant peptide deformylase, an essential enzyme with potential for the development of novel biotechnology applications in agriculture. Biochem. J. **413**:417-427.

10. **Dirk, L. M. A., M. A. Williams, and R. L. Houtz**. 2001. Eukaryotic peptide deformylases. Nuclear-encoded and chloroplast-targeted enzymes in arabidopsis. Plant Physiol. **127**:97-107.

11. **Escobar-Alvarez, S., Y. Goldgur, G. Yang, O. Ouerfelli, Y. Li, and D. A. Scheinberg**. 2010. Structure and activity of human mitochondrial peptide deformylase, a novel cancer Target. Mol. Cell. Biol. **30**:5099-5109

12. **Ferrer, M., F. Martınez-Abarca, and P. N. Golyshin**. 2005. Mining genomes and `metagenomes' for novel catalysts. Curr. Opin. Biotechnol. **16**:588-593.

13. **Giglione C, .** 2000. Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. Mol. Microbiol**. 36**:1197-1205.

14. **Giglione, C.** 2000. Identification of eukaryotic peptide deformylases reveals universitality of N-terminal protein processing mechanisms. EMBO J. **19**:5916-5929.

15. **Groche, D., A. Becker, I. Schlichting, W. Kabsch, S. Schultz, and A. F. V. Wagner**. 1998. Isolation and crystallization of functionally competent *Escherichia coli* peptide deformylase forms containing either iron or nickel in the active site. Biochem. Biophys. Res. Comm. **246**:342-346.

16. **Guilloteau, J. P., M. Mathieu, C. Giglione, V. Blanc, A. Dupuy, M. Chevrier, P. Gil, A. Famechon, T. Meinnel, and V. Mikol**. 2002. The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: a platform for the structure-based design of antibacterial agents. J. Mol. Biol. **320**:951-962.

17. **Hu, Y. J., Y. Wei, Y. Zhou, P. T. R. Rajagopalan, and D. Pei**. 1999. Determination of substrate specificity for peptide deformylase through the screening of a combinatorial peptide library. Biochemistry **38**:643-650.

18. **Kreusch, A., G. Spraggon, C. C. Lee, H. Klock, D. McMullan, K. Ng, T. Shin, J. Vincent, I. Warner, C. Ericson, and S. A. Lesley**. 2003. Structure analysis of peptide

deformylases from *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Thermotoga maritima* and *Pseudomonas aeruginosa*: snapshots of the oxygen sensitivity of peptide deformylase. J. Mol. Biol. **330**:309-321.

19. **Lang, B. F., M. W. Gray, and G. Burger**. 1999. Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet. **33**:351-397.
20. **Leeds, J. A. and C. R. Dean**. 2006. Peptide deformylase as an antibacterial target: a critical assessment. Curr. Opin. Pharmac. **6**:445-452.
21. **Li, J. Y., Y. M. Cui, L. L. Chen, M. Gu, J. Li, F. J. Nan, and Q. Z. Ye**. 2004. Mutations at the S1 sites of methionine aminopeptidases from *Escherichia coli* and *Homo sapiens* reveal the residues critical for substrate specificity. J. Biol. Chem. **279**:21128-21134.
22. **Mannervik, B., A. Runarsdottir, and S. Kurtovic**. 2009. Multi-substrate–coactivity space and quasi-species in enzyme evolution: Ohno's dilemma, promiscuity and functional orthogonality. Biochem. Society Transactions **037**:740-744.
23. **Margolis, P., C. Hackbarth, S. Lopez, M. Maniar, W. Wang, Z. Yuan, R. White, and J. Trias**. 2001. Resistance of *Streptococcus pneumoniae* to deformylase inhibitors is due to mutations in *def*B. Antimicrob. Agents Chemother. **45**:2432-2435.
24. **Mazel, D., S. Pochets, and P. Marliere**. 1994. Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO **13**:914-923.
25. **Meinnel, T. and S. Blanquet**. 1995. Enzymatic properties of *Escherichia coli* peptide deformylase. J. Bacteriol. **177**:1883-1887.
26. **Meinnel, T., Y. Mechulam, and S. Blanquet**. 1993. Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli*. Biochimie **75**:1061-1075.
27. **Narita, M., S. Honda, H. Umeyama, and S. Obana**. 1988. The solubility of peptide intermediates in organic solvents. Solubilizing potential of hexafluoro-2-propanol. Bull. Chem. Soc. Jpn. **61**:281-284.
28. **Nguyen, K. T., X. Hu, C. Colton, R. Chakrabarti, M. X. Zhu, and D. Pei**. 2003. Characterization of a human peptide deformylase: implications for antibacterial drug design. Biochemistry **42**:9952-9958.
29. **Owusu, R. K. and D. A. Cowan**. 1989. Correlation between microbial protein thermostability and resistance to denaturation in aqueous: organic solvent two-phase systems. Enzyme Microbial. Technol. **11**:568-574.
30. **Peters, R., J. Hellenbrand, Y. Mengerink, and S. Van der Wal**. 2004. On-line determination of carboxylic acids, aldehydes and ketones by high-performance liquid chromatography-diode array detection-atmospheric pressure chemical ionisation mass spectrometry after derivatization with 2-nitrophenylhydrazine. J. Chromatogr. **1031**:35-50.
31. **Pichota, A., J. Duraiswamy, Z. Yin, T. H. Keller, J. Alam, S. Liung, G. Lee, M. Ding, G. Wang, W. L. Chan, M. Schreiber, I. Ma, D. Beer, X. Ngew, K. Mukherjee, M. Nanjundappa, J. W. P. Teo, P. Thayalan, A. Yap, T. Dick, W. Meng, M. Xu, J. Koehn, S. H. Pan, K. Clark, X. Xie, C. Shoen, and M. Cynamon**. 2008. Peptide deformylase inhibitors of Mycobacterium tuberculosis: synthesis, structural investigations, and biological results. Bioorg. Med. Chem. **18**:6568-6572.
32. **Ragusa, S., P. Mouchet, C. Lazennec, V. Dive, and T. Meinnel**. 1999. Substrate recognition and selectivity of peptide deformylase. similarities and differences with metzincins and thermolysin. J. Mol. Biol.**289**:1445-1457.
33. **Rajagopalan, P. T. and D. Pei**. 1998. Oxygen-mediated inactivation of peptide deformylase. J. Biol. Chem. **273**:22305-22310.

34. **Serero, A., C. Giglione, and T. Meinnel**. 2001. Distinctive features of the two classes of eukaryotic peptide deformylases. J. Mol. Biol. **314**:695-708.

35. **Serero, A., C. Giglione, A. Sardini, J. Martinez-Sanz, and T. Meinnel**. 2003. An unusual peptide deformylase features in the human mitochondrial *N*-terminal methionine excision pathway. J. Biol. Chem. **278**:52953-52963.

36. **Sonke, T., B. Kaptein, A. F. V. Wagner, P. J. L. M. Quaedflieg, S. Schultz, S. Ernste, A. Schepers, J. H. M. Mommers, and Q. B. Broxterman**. 2004. Peptide deformylase as biocatalyst for the synthesis of enantiomerically pure amino acid derivatives. J. Mol. Catal. B: Enzymatic **29**:265-277.

37. **Walker, K. W. and R. A. Bradshaw**. 1999. Yeast methionine aminopeptidase I. J. Biol. Chem. **274**:13403-13409.

Chapter 4

Screening and selection methods for deformylase activity

Abstract

Several screening and selection methods were tested for their application in the screening of directed evolution libraries of *E. coli* peptide deformylase. In microtiter plate format, activity could be detected using a pH indicator that changes color due to formic acid formation. For selection assays, we optimized a protocol based on complementation of L-leucine auxotrophy of host cells due to peptide deformylase mediated production of a deformylated precursor of L-leucine. The method was used to select active clones on synthetic medium containing arabinose for induction and glucose and amino acids to suppress toxic effects caused by overexpression of peptide deformylase.

Introduction

The last decade has seen an increasing number of applications of enzymes in various industrial sectors, especially because of the high selectivity and activity of many enzymes under mild reaction conditions. On the other hand, many enzymes are not ideal process catalysts [14]. Over the last two decades, the growing understanding of structure-function relationship in proteins and the development of efficient molecular biology protocols have increased the possibilities for studying and engineering enzyme properties. Thus, protein engineering has become a powerful tool for improving the fitness of enzymes for various practical applications [7 1;39]. If directed evolution is used, a success will not only require the choice of a good method for library construction, but also the availability of a reliable way to indentify and isolate the desired variants from large collections of mutant enzymes [19;40]. Indeed, the development of analytical tools for finding the improved variants (sometimes compared to searching for a needle in the haystack) often is the bottleneck during the improvement of an industrial enzyme by directed evolution [29].

For retrieving the best enzyme variants from a mutant library, it is essential that the desired feature is linked to the gene encoding this particular variant. This phenytype-genotype linkage can be achieved in various ways, such as phage or ribosome display [11;27;32], cell surface display [3], whole-cell methods, or through in vitro compartmentalization [13].

Identification of the best variants (phenotypes) in a mutant library can be achieved in two fundamentally different ways: screening or selection. Although the two terms are often confused in the literature, an important difference is that in case of screening the mutants composing the library are tested individually via a suitable analytical method, whereas selection is based on physical (affinity-based panning) or biological (using a specific growth advantage) purification or enrichment of the desired variants from the mutant library. For the latter, the genotype-coupled improved enzyme must make the system selectable, e.g. by enhancing the affinity for a specific ligand or by providing to the host the capacity to survive or multiply [34;38]. During the design of a screening or selection protocol, it should be kept in mind that "you get what you screen for". This means that it is highly desirable to use screening or selection conditions that are as closely as possible related to the enzyme property that needs to be improved for a real application [41]. Consequently, cost effective, reproducible, and rapid high-throughput techniques are of major importance.

A range of screening and selection strategies have been described [9;10;21;22;26;34] and choosing the best one is not always an easy task. Often, a two-round procedure is used, in which a first round of screening or selection is designed to discard all deleterious mutants, and a second round, which is carried out under more stringent conditions with the real substrate, is used to recover the best variants for a specific task [5]. The initial screening round is often simplified by coupling the formation of the product of the desired reaction to a follow-up transformation that results in a visible signal, or by employing artificial substrates that lead to a coloured or fluorescent product that can easily be detected [31 12;41]. However the use of artificial substrates bears the inherent risk that a biocatalyst identified does not have the required features in the real conversion [34].

The advantage of screening is that it is based on the separate examination of each variant in the library for the desired characteristic (e.g. increased activity) via an analytical technique like liquid chromatography, mass spectrometry, or NMR. This allows a direct measurement of product and substrate, providing quantitative data, and conditions may be chosen that reflect, in miniaturized form, the real application [36]. Clearly, depending on the size of the library, screening can become very laborious and expensive, also because many mutants are tested that are not active or not properly folded. Costs can also increase since the design of appropriate small-volume assays often involves robotics (colony pickers, automatic liquid handlers) and sophisticated analytical equipment like MS or NMR instruments equipped with flow injection, or

fluorescence-activated cell sorting (FACS) [5;34;36]. Furthermore, even at microtiter plate scale large numbers of assays may add to the cost of the procedure when expensive chemicals such as cofactors are needed.

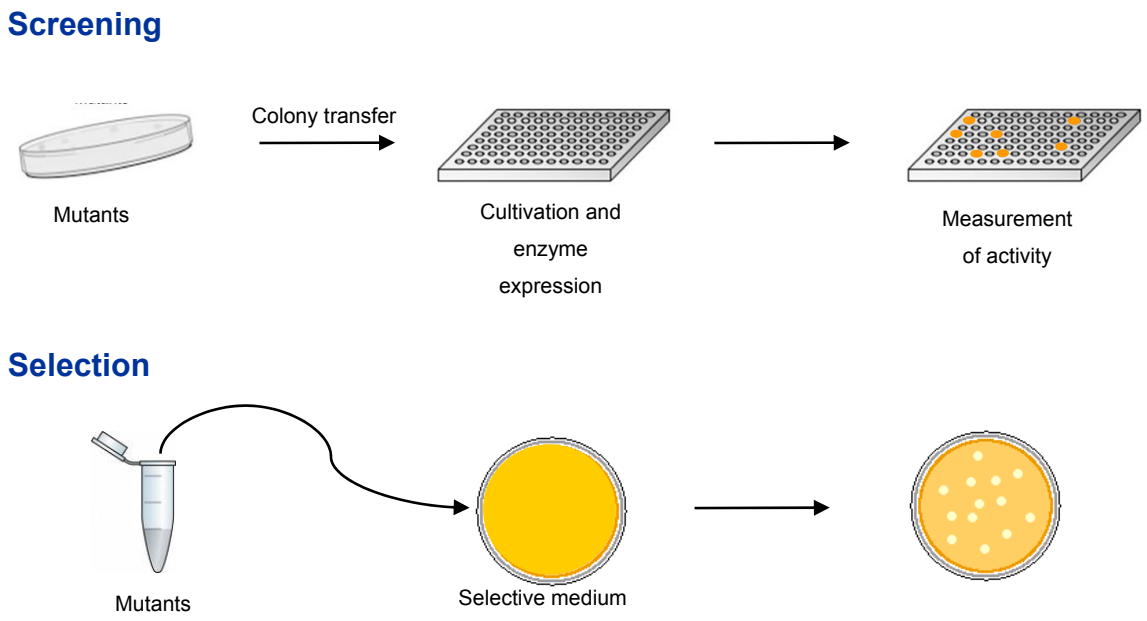


Figure 1. Visualization of differences between screening and selection.

These drawbacks of screening stand in strong contrast to the use of growth selection strategies, which are much less labour intensive and allow analysis of much larger numbers of variants at the same time. Library sizes of up to 10^{10}-10^{13} variants can be handled by selection, compared to some 10,000, which is often considered the upper limit for most microtiter plate-based screenings [8]. Careful design of a selection assay is crucial to prevent false positives that could appear, for example as a result of chromosomal mutations that influence the selected phenotype.

The objective of this study was to develop a method for the isolation of peptide deformylase (PDF) variants with increased activity toward selected formylated dipeptides, starting from a library composed of about 10^6 variants that was obtained by simultaneous saturation of 4 positions in the enzyme sequence (Chapter 5). We describe various screening methods that were investigated as well as the development of a selection method based on the complementation of *E. coli* leucine auxotrophy.

Results and discussion

Formate dehydrogenase assay

The most commonly used assay for PDF is the one developed by the group of Meinnel [20], which is based on coupling the production of formate to the reduction of NAD^+ by formate dehydrogenase (FDH). The deformylation rate is directly proportional to the rate of production of NADH from NAD^+, which can be monitored at 340 nm. Unfortunately, the low specificity constant of FDH for formate can easily make formate conversion the rate-limiting step in this assay, corrupting a correct evaluation of PDF activity. Moreover, the high background absorbance at 340 nm before the addition of substrate to the reaction mixture containing cell-free extract will also interfere with a correct assessment of the deformylation activity.

We did some tests but it appeared that the assay indeed was too insensitive to reliably distinguish clones converting *N*-formyl-Met-Ala-OH from background (cell free extract not overexpressing EcPDF). The results did not match those obtained with the standard HPLC test, and the formate dehydrogenase-based assay for library screening was abandoned.

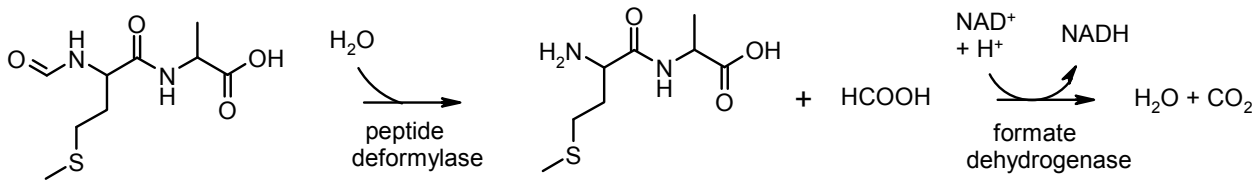


Figure 2. Reaction scheme of the formate dehydrogenase assay for peptide deformylase with *N*-formyl-Met-Ala-OH as substrate.

Ninhydrin method

The ninhydrin method is based on a chromogenic reaction that detects the primary amine group of methionine which becomes unmasked in the deformylation reaction. First, the unspecific background reaction of ninhydrin with free amines present in the *E. coli* crude extracts was determined. Different dilutions of this extract were incubated with a ninhydrin solution at 90°C for 5 and 10 min. The results showed that only at total protein concentrations below 1 mg/mL the samples did not show the typical Ruhemann's purple colour. Subsequently, extracts prepared from *E. coli* cells overexpressing peptide deformylase from plasmid pBAD-PDF_{Etag} (Chapter 2) were diluted to a total protein concentrations of about 0.1 mg/mL and incubated with 10 mM

N-formyl-Met-Ala-OH or *N*-formyl-Phe-Phe-OH. From assays with extracts containing different levels of PDF, it became apparent that very precise dilutions of cell-free extract and exact incubation times were required to allow discrimination between the wild-type enzyme and improved mutants. Furthermore, it was judged difficult to sufficiently control these variables when the assay was done in high-throughput format. Moreover, peptides with a primary amino group, e.g. *N*-formyl-Lys-Phe-OH, can react with ninhydrin in the absence of deformylation, disturbing the assay. These facts, especially the low sensitivity and high background signals, indicated that a more robust method was needed for library screening.

p-(Dimethylamino)benzaldehyde (DMBA) method

A second colorimetric assay that was tested is based on p-(dimethyl-amino)benzaldehyde (DMBA). DMBA reacts rapidly with primary amines to form a complex with a maximum absorbance at 450 nm which is easily detectable by spectro-photometry (Fig.3). A linear relation between absorbance and peptide concentration was obtained in the range of 0.3-10 mM H-Met-Ala-OH. Furthermore, it was possible to detect the difference in deformylase activity between a clone overexpressing $EcPDF_{Etag}$ and one in which only the genomic copy of the PDF-encoding gene was present. The activities on *N*-formyl-Met-Ala-OH that were measured with the DMBA assay correlated well to the ones obtained via HPLC analysis. Thus the DBMA assay seems applicable for high-throughput screening of mutant libraries.

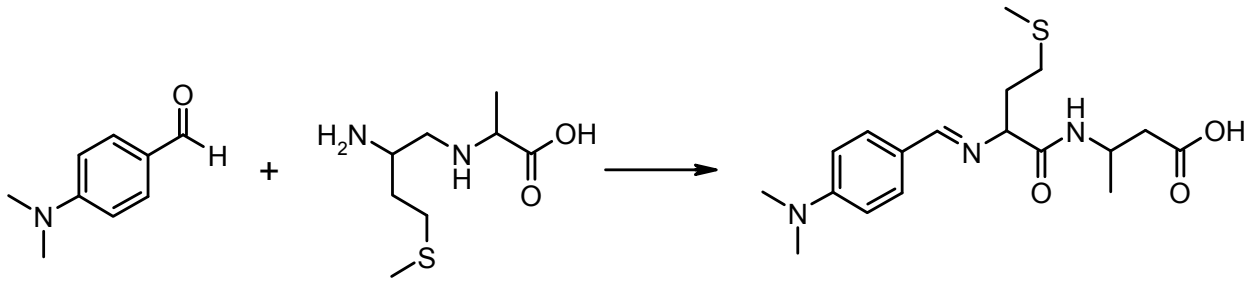


Figure 3. Reaction of DMBA with H-Met-Ala-OH, thereaction product of the PDF-catalyzed deformylation of *N*-formyl-Met-Ala-OH.

Despite the positive results obtained, it was decided not to use this method for primary screening of PDF variants. The robotics needed for the processing a library composed of 10^6 clones makes the method too expensive.

pH shift method

Aiming at a screening method that would allow an easy and fast examination of a large number of colonies, we attempted the use of pH indicators. Advantages are easy detection of colour change without the need for a chromogenic moiety present on the substrate, and the possibility to use an indicator directly on agar plates [6;34].

During the deformylation reaction catalyzed by PDF, formic acid (pK_a=3.75) and a free N-terminal amino group (pK_a=8.0) are formed. The initial pH will decrease during progress of the reaction, which can be detected by a pH indicator incorporated in an agar plate or added to a microtiter plate. First, we investigated the decrease in pH associated with the deformylation reaction by incubating about 50 mg of *E. coli* cells that did or did not overexpress $EcPDF_{Etag}$ with a buffered substrate solution containing *N*-formyl-Met-Ala-OH (2.5 mM) and phenol red. The pH of the solution was 7.2. When using a phosphate buffer concentration of 20 mM, the pH decreased 1.2 units in 5 min, whereas this was only 0.5 units in case of the blank reaction (*E. coli* cells, not induced).

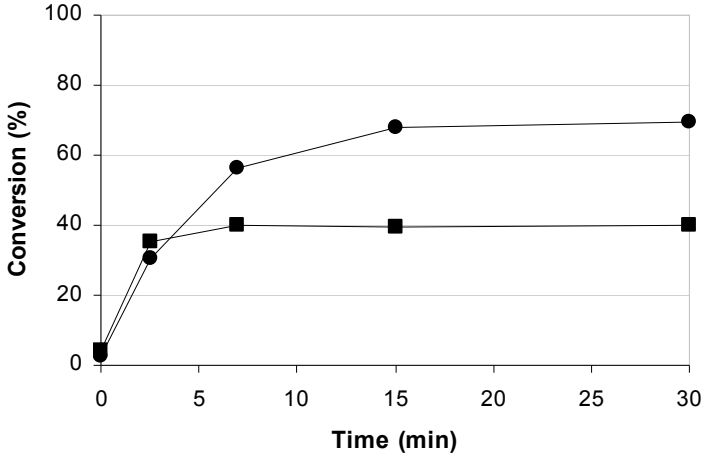


Figure 4. Effect of pH indicator on the activity of $EcPDF_{Etag}$ with N-formyl-Met-Ala-OH. Symbols: ●, reaction without phenol red; ■, reaction in the presence of phenol red.

Subsequently, the effect of the pH indicator on the activity of $EcPDF_{Etag}$ was tested. The decrease of pH was followed in a similar reaction mixture with and without addition of phenol red. This yielded a pH decrease of 1.6 units in 5 min in both cases. However, when the amount of *N*-formyl-Met-Ala-OH and H-Met-Ala-OH were analyzed using the standard HPLC assay, 70 % substrate conversion appeared to have been reached when no phenol red was added to the reaction mixture, while only 40 % conversion was observed under the same conditions in the presence of the indicator (Fig. 4). Apparently, this difference in conversion caused only a marginal effect on the in

pH shift. Even though the inhibitory effect of the indicator was only small when using *N*-formyl-Met-Ala-OH as the substrate, it was decided not to continue the development of this pH indicator based method. Careful tuning of the buffer concentration for each specific screening substrate (especially for the poorly converted ones), and of reaction time and amount of PDF to be used in the screening assay would have been needed to make this screening method sufficiently sensitive and discriminating.

Growth selection

Because of the very high throughput and the minimal equipment needed, selection methods were explored. To allow a link between the growth of *E. coli* and the deformylase activity, we used leucine-containing *N*-formylated dipeptides as sole source of leucine to complement the leucine auxotrophy of *E. coli* strains like MC1061 and TOP10. The following requirements hold for such a selection method: (1), *E. coli* auxotrophic strains should be able to grow on a minimal medium containing leucine; (2), the cells need to take up the leucine-containing formylated peptide from the medium; (3), the desired EcPDF variants have to deformylate the *N*-formylated dipeptide; (4), cell-born peptidases and/or amidases need to hydrolyze the deformylated dipeptide, and not the formylated one, liberating leucine; and (5), leucine, that is produced this way should sustain growth.

A selection assay was explored by Mazel et al. [24] to identify and isolate the deformylase gene of *E. coli*. The selection was based on the ability of methionine and leucine auxotrophic cells, transformed with indigenous *E. coli* genes, to hydrolyse *N*-formyl-Leu-Met-OH. This compound can not be degraded by wild-type *E. coli* cells but cell growth was achieved through overexpression of *E. coli* DNA fragments, cloned into a high-copy-number plasmid, conferring to the cell the ability to use the deformylase substrate as a source of methionine and leucine. This method shows that plate selection of PDF variants is possible. A similar strategy was explored by us for the selection of PDF mutants with altered P1 specificity, using a leucine auxotrophic *E. coli* strain as a host, overexpression of EcPDF, and selecting of variants on a medium containing *N*-formylated leucine dipeptides as sole source of leucine.

For a successful selection assay, the medium has to be designed in such a way that non specific-growth (by false-positives) is avoided. The choice of a suitable minimal medium started by testing the ability of *E. coli* MC1061 and TOP10 cells transformed with pBAD-PDF_{Etag} to grow on M9 minimal medium containing glycerol as carbon

source. This medium also contained thiamine, carbenicillin, 0.002% (w/v) L-arabinose to induce the PDF expression, and when needed a source of leucine (as described in Materials and Methods). Growth tests showed that the presence of L-leucine promoted the formation of colonies after 3 days, whereas without leucine no colonies appeared. The same experiment was carried out using the dipeptide *N*-formyl-Asp-Leu-OH as sole leucine source, but in this case colonies were not visible after 5-15 days of incubation at 37°C. To demonstrate that this result was not caused by toxicity of the substrate, the viability of the cells was checked by applying small filter paper discs (diameter ca. 0.5 cm) with a 300 µM L-leucine solution to the agar plates. The formation of colonies around the filter after 1-2 days of incubation confirmed that cells were still viable. After incubation for 15 days at 37°C on the selection medium the *E. coli* cells thus are still viable and unable to utilize *N*-formyl-Asp-Leu-OH as leucine source, indicating that the selection medium is sufficiently stringent to prevent false positives. The lack of growth with *N*-formyl-Asp-Leu-OH is in agreement with the very low activity of EcPDF towards formylated dipeptides with aspartic acid in position P1' [30]. The test revealed that, under the applied conditions, discrimination between wild-type EcPDF and a mutant with increased activity towards *N*-formyl-Asp-Leu-OH should be possible.

After showing that wild-type cells were not able to grow on selection medium, a positive control demonstrating the uptake and metabolism of formylated dipeptides was necessary. Four shake flasks containing liquid M9 minimal medium, with glycerol and arabinose, and with *N*-formyl-Asp-Leu-OH, H-Asp-Leu-OH, *N*-formyl-Leu-Phe-NH_2 or H-Leu-Phe-NH_2 as sole leucine source, were inoculated with *E. coli* MC1061 containing pBAD-PDF_{Etag}. Growth rate and dipeptide concentration were monitored to demonstrate uptake and subsequent peptide bond hydrolysis (Fig. 5). Both non-formylated dipeptides supported growth of the cells, which coincided with dipeptide consumption. The two formylated dipeptides, on the other hand, were not consumed at all, and did not enable the growth of *E. coli* cells. The lack of growth with the formylated dipeptides confirms the low activity of peptide deformylase with the substrates used in these experiments. Next, we tested *N*-formyl-Met-Leu-OH, which is a good PDF substrate,

A B

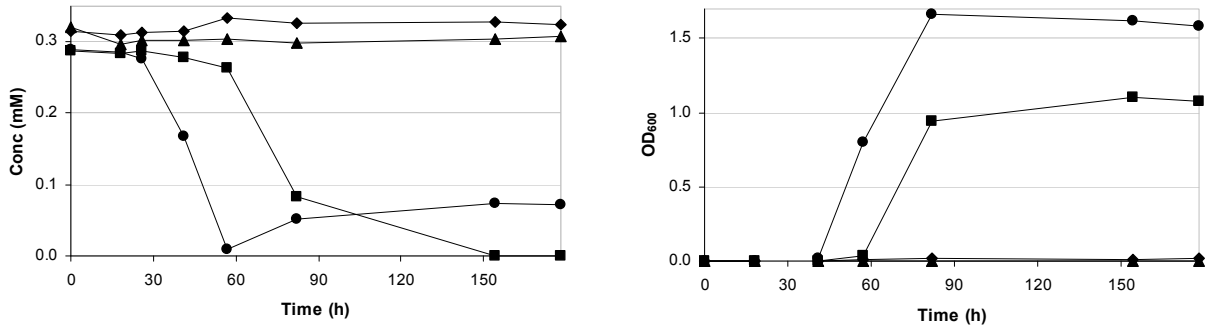


Figure 5. Growth and dipeptide uptake by *E. coli* MC1061 cells containing pBAD-PDF_{Etag}. (A) Consumption of dipeptides. (B) Growth in the presence of different dipeptides as L-leucine source. Symbols: ▲, *N*-formyl-Leu-Phe-NH_2; ♦, *N*-formyl-Asp-Leu-OH; ●, H-Leu-Phe-NH_2; ■, H-Asp-Leu-OH.

as source of leucine. Ragusa and coworkers observed that the best substrates for PDF are formylated dipeptides with methionine in position P1' [30]. We tested the growth of *E. coli* strains MC1061 and TOP10 on M9 minimal medium plates supplied with *N*-formyl-Met-Leu-OH, in the absence of L-arabinose. Colonies appeared after 2 days at 37°C, showing that the cells were able to take up and metabolize the substrate via deformylation by their genome encoded peptide deformylase and subsequent peptide bond hydrolysis by endogenic peptidases. The negative control in which no peptide was added to the minimal medium showed no growth. The experiment was repeated with recombinant cells overexpressing $EcPDF_{Etag}$ from a pBAD plasmid and induced with increasing concentrations of L-arabinose (0.002 to 0.2% w/v). Interestingly, we noticed a decrease of the number and the size of colonies, which was proportional to the inducer concentration. Using M9 medium with 0.02% L-arabinose, the number of CFU decreased to 4% of a control on M9 medium with no inducer. At concentrations higher than 0.02 % (w/v) L-arabinose, no growth was visible even after several days. When the same strains were plated on LB medium with varying concentrations of arabinose, we found that growth was not influenced by induction.

Tao and coworkers [37] showed that gene expression is different in *E. coli* cells growing in LB medium compared to cells cultivated in minimal media. In the latter case, the need for building blocks switches on biosynthetic pathways, including regulators of metabolism and stress response factors. Cell growth is slower and metabolites (e.g. acetate) accumulate in the cells. A study showing that the expression of proteins from vectors containing the pBAD promoter varies in different media was published by Guzman and colleagues[15]. Their results showed more efficient expression of

heterologous genes in glycerol containing minimal medium compared to rich medium, which could be explained by partial catabolite repression or specific repression by components present in the rich medium. From these observations we expected that growth on glycerol-containing minimal medium would be associated with strongly induced peptide deformylase expression, which could cause a level of stress to the cells that is not compatible with healthy growth. To check this, *E. coli* MC1061 and TOP10 empty strains (expressing no PDF) were plated on minimal medium containing different L-arabinose concentrations. The ability to grow even at 0.2 % (w/v) L-arabinose confirmed that induced overexpression, likely in combination with activation of biosynthetic pathways, negatively influenced growth on synthetic medium.

We therefore decided to test if the activity of PDF or the PDF protein itself impaired growth. Inactive PDF mutants were created by site-directed mutagenesis of three conserved residues reported to be involved in the catalytic cycle [25]. Two mutants of $EcPDF_{Etag}$ were constructed: (i) C138A/H141A and (ii) E142A. Overexpression and activity tests resulted in two soluble mutant proteins lacking deformylase activity. Both mutant plasmids were transformed to *E. coli* MC1061 and TOP10. After plating on glycerol minimal medium supplemented with 0.002 or 0.2 % (v/w) L-arabinose, visible colonies only appeared on plates containing the lower concentration of inducer, clearly showing that the negative effect of induction on growth is not caused by an excess of deformylase activity in the cells but by the high activity of the transcription/translation machinery under the strongly inducing conditions, or by the PDF protein itself. Note that peptide deformylase is not just an ordinary cytoplasmic enzyme. Recently Bingel-Erlenmeyer and coworkers [4] showed that there is a specific interaction of EcPDF with the ribosome. This association, mediated by the C-terminal helix, is important *in vivo* to ensure efficient processing of newly synthesised peptides. Since ribosomal peptide synthesis is highly organized and concerted, it is conceivable that an unbalance in the relative levels of ribosome and PDF in the cells negatively effects translation under conditions of stress imposed by minimal medium and high overexpression.

To find a minimal medium that could support growth at high L-arabinose concentrations, we tested the addition of nutrients to minimal medium. We first tested casamino acids, which is a peptide mixture produced by hydrolysis of casein. Separate samples containing equal numbers of colony forming units (CFU) of *E. coli* TOP10(pBAD-PDF_{Etag}) or MC1061(pBAD-PDF_{Etag}) were plated and incubated on rich medium (LB) and minimal medium (supplemented with glycerol, thiamine, $MgSO_4$,

$CaCl_2$ and casamino acids). In both cases increasing concentrations of L-arabinose were tested (0 - 0.002 - 0.02 or 0.2 %). The number of appearing CFU was higher when casamino acids were present in the minimal medium, but growth was still not visible at 0.2-0.02% (w/v) L-arabinose. Also when trace elements (TEs) were added to the minimal medium, the growth inhibiting effect of the inducer was still occurring. Only when both casamino acids and TEs were added, colonies appeared (in limited number) after incubation at 37°C for 5 days, and 90% of the plated CFUs formed colonies on synthetic medium containing casamino acids and TEs at 0.002% L-arabinose.

Another parameter that is crucial for growth and gene expression is the carbon source. The growth rate of *E. coli* is about 2-fold higher in glucose minimal medium than with glycerol, but the total level of gene transcription is higher in case of a carbon source that is more rapidly utilized [23;28]. The cells not only upregulate genes involved in the metabolism of the given carbon source, but also activate pathways linked to unavailable compounds. This broader transcription program, although requiring more energy, ensures the cells a faster stress response in nutrient-poor environments [2;16;23].

To further decrease the stress applied on the cells, glycerol as carbon source in the minimal medium was replaced by glucose [42]. The expression system used in this study (pBAD vector) is based on the L-arabinose operon. This system is characterized by the possibility to modulate expression rates by increasing or decreasing the inducer concentration, and a very low level of uninduced expression, which can be reduced to zero by addition of glucose The influence of glucose is mediated by cAMP, which binds CRP (cAMP receptor protein), which together with AraC is necessary for induction [17;33]. Moreover glucose can inhibit L-arabinose uptake by repressing the synthesis of L-arabinose permease [18]. The use of glucose in combination with TEs and casamino acids efficiently minimized the influence of the L-arabinose inducer concentration on cell growth. CFU counts were comparable in LB and M9 medium containing TEs, casamino acids and glucose even at 0.2% (w/v) L-arabinose.

These observations show that a combination of TEs, casamino acids and glucose could overcome the growth-inhibitory effect of PDF induction. However, since the casamino acid mixture contains a source of leucine, it can not be used in the current selection procedure. Therefore, a mixture of amino acids in which leucine was not included was prepared as described in the Experimental section. Even though a selection method using this medium containing both glucose and L-arabinose seems contra-intuitive, overexpression of PDF under these conditions was observed on SDS-

polyacrylamide gels using CFE prepared from cells grown on glucose medium. Although clone to clone variations on the same plate were observed at L-arabinose concentrations of 0.2-0.02%, a clear expression of peptide deformylase was detected, confirming that the presence of glucose is not leading to complete inhibition of PDF expression.

Finally, different concentrations of dipeptides (H-Leu-Phe-OH, H-Asp-Leu-OH and H-Phe-Leu-OH) were added as leucine source to the glucose medium containing TEs and amino acids to determine the optimal dipeptide concentration for growth of *E. coli* when the *N*-formylated dipeptides mentioned below would be used as selection substrates. This was found to be 1 mM. In addition no chemical hydrolysis was detected by HPLC analysis after 2 weeks of incubation at 37°C in minimal medium for none of the tested substrates (*N*-formyl-Phe-Leu-OH, *N*-formyl-Leu-Phe-NH_2, *N*-formyl-Asp-Leu-OH). Thus, it was expected that PDF variants with suitably altered specificity in the P1' pocket can be isolated by growth selection on glucose-containing mimimal medium supplemented with a mixture of amino acids and 1 mM of a leucine-containing formylated di- or tripeptide as sole source of leucine. Because of these positive results and the ease of use, with no background growth, growth selection on agar plates was concluded to be the most attractive method for the identification of PDF mutants with increased activity towards synthetic peptide substrates.

Conclusions

The development of robust and cost effective methods for assaying enzyme activity is extremely important for the identification of desired enzyme variants by protein engineering. In this study several screening methods were compared and a selective growth method was fine-tuned for the selection of modified peptide deformylases in *E. coli* leucine auxotrophic cells. The use of glucose as carbon source allowed growth of PDF overexpressing variants, providing a tool for selection of PDFs with modified activity from directed evolution libraries.

Experimental section

Materials

All standard chemicals were of the highest grade obtainable. *N*-formyl-Met-Ala-OH, H-Met-Ala-OH, *N*-formyl-Met-Lys-OH, *N*-formyl-Leu-Phe-NH_2, H-Leu-Phe-OH, H-Phe-Leu-OH, H-Asp-Leu-OH, were purchased from Bachem (Bubendorf, Switzerland). Bovine liver catalase (10,000-40.000 U/mg) was obtained from Sigma-Aldrich (Saint Louis, USA); tris-(2-carboxyethyl)-phosphine (TCEP) was obtained from Fluka (Buchs, Switzerland). Formylated amino acid derivatives were obtained from Chiralix (Nijmegen, NL). Amino acids were purchased from Sigma-Aldrich.

Restriction enzymes, T4 DNA ligase and T4 DNA polymerase were obtained from New England Biolabs. Adenosine triphosphate and deoxynucleoside triphosphates were purchased from Roche (Mannheim, Germany). Synthesis of oligonucleotides for site-directed mutagenesis and DNA sequencing were performed at Life Technologies GIBCO (Carlsbad, USA).

Microorganism, plasmids and cultivation conditions in growth assays

Plasmid-free E. coli TOP10 (mcrA, Δ(mrr-hsdRMS-mcrBC), ΔlacX74, deoR, recA1, araD139Δ(ara-leu)7697, galK, rpsL, endA1, nupG) and MC1061 (araD139 Δ(araA-leu)7697 ΔlacX74 galK16 galE15(GalS) λ- e14- mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2) were propagated in Luria Bertani (LB) medium at 37°C. When cells were transformed with pBAD-PDFwt or pBAD-PDFEtag, 100 mg/L carbenicillin was added to the medium. For the growth tests, E. coli cells were plated on different M9 based minimal media and incubated at 37°C.

The M9 minimal medium (pH 7.3) contained per liter: 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 240 mg $MgSO_4$ and 11 mg $CaCl_2$. When necessary 10 mg/L thiamine, and/or 100 mg/l of cabenicillin were added. Glucose (0.2 % w/v) or glycerol (0.2 % w/v)) were added as carbon source. Solid medium contained 20 g/l agarose. When positive control tests were run, 0.4 mM leucine was added while if the substrate used was a formylated dipeptide, its final concentration was 1 mM. L-arabinose concentrations varied form 0 to 0.2% (v/w).

Casamino acids M9 medium was prepared by addition of casamino acids (Becton Dickinson, USA) to M9 medium at a final concentration of 10 mg/L. In this case no L-Leu was added.

When required, amino acids (50 mg L-alanine, 50 mg L-arginine, 500 mg L-aspartic acid, 500 mg L-glutamic acid, 50 mg glycine, 50 mg L-histidine, 50 mg L-isoleucine, 50 mg L-lysine, 50 mg L-methionine, 50 mg L-phenylalanine, 50 mg L-proline, 50 mg L-serine, 50 mg L-threonine, 20 mg L-tryptophan, 20 mg L-tyrosine and 50 mg L-valine), and 5 mL of a trace element stock solution were added to M9 medium (1 liter). Trace elements were prepared as a 1 L stock containing the following: 780 mg $Ca(NO_3)_2 \cdot 4H_2O$, 200 mg $FeSO_4 \cdot 7H_2O$, 10 mg $ZnSO_4 \cdot 7H2O$, 10 mg H_3BO_3, 10 mg $CoCl_2 \cdot 6H_2O$, 10 mg $CuSO_4 \cdot 5H_2O$, 4 mg $MnSO_4 \cdot H_2O$, 3 mg $Na_2MoO_4 \cdot 2H_2O$ and 2 mg $NiCl_2 \cdot 6H_2O$.

Electroporation of *E. coli* MC1061 cells

E. coli MC1061 cells were made electrocompetent by inoculating a single colony into 5 mL LB medium, followed by overnight growth at 28°C. Next, 2.5 mL of this culture was transferred to 500 mL LB medium and grown at 37°C to an OD_{620} of ca. 0.5. Cells were then harvested by centrifugation for 20 min at 4,400x g in a Beckman JLA 8.1000 rotor and kept at 4°C for subsequent steps. The pellet was resuspended in 500 mL ice-cold water and centrifuged again. This procedure was repeated twice. The pellet volume was estimated and an equal volume of ice-cold water was added to resuspend cells. This yielded electrocompetent cells that could be stored at -20°C.

For electroporation, 0.2 ng of plasmid DNA was added to 50 μL of electrocompetent cells followed by application of a pulse in an electroporation cuvette (400 ohms, 2.5 kV, 25 μF). After addition of 1 mL of SOC medium, the cells were incubated at 37°C for one h before plating appropriate dilutions on LB plates supplemented with 100 mg/L carbenicillin.

Protein analysis

Protein concentrations were determined using the Bradford method with bovine serum albumin (BSA, Sigma) as standard protein. [35]

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was routinely performed in NuPAGE 4 to 12% Bis-Tris gels (Invitrogen) under reducing conditions with MOPS-SDS running buffer. The gels were stained with the SimplyBlue SafeStain (Invitrogen).

Formate dehydrogenase assay for peptide deformylase

The FDH method is based on the use of Candida boidinii formate dehydrogenase. In a quartz cuvette (1 cm optical path), 450 µL reaction mixture containing 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase, 10 mM NAD^+ solution, 1 mM N-formyl-Met-Ala-OH and 1U FDH (Sigma) were incubated. The reaction was started by adding 50 µL of a diluted EcPDF solution in 20 mM MOPS/NaOH, pH 7.7, 100 mM NaCl, 10 µg/mL bovine liver catalase, and 2 mM TCEP to the mixture. The reaction was monitored at 30°C using a Perkin-Elmer UV-Vis spectrophotometer at 340 nm.

Ninhydrin assay

The ninhydrin assay for the detection of deformylase activity was carried out at 30°C in a total volume of 200 µL. An amount of 100 µL of a suitable dilution of PDF in a buffer containing 20 mM MOPS/NaOH, 10 µg/mL bovine liver catalase, 100 mM NaCl, pH 7.7, and 2 mM TCEP was mixed with 100 µL of substrate solution composed of 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase (Sigma) and 10 mM of the substrate (*N*-formyl-Met-Ala-OH or *N*-formyl-Phe-Phe-OH). The PDF dilutions were chosen such that approximately 10% of the substrate was converted in 15 min. Reactions were stopped by addition of 100 µL of a ninhydrin solution (1% in H_2O) followed by incubation at 90°C for 5 min.

DMBA assay

The p-dimethylaminobenzaldehyde (DMBA) reagent was prepared by dissolving 2 g DMBA in a citric acid solution in ethanol (315.5 g citric acid/L ethanol). The deformylase reaction was conducted in the same buffer as the ninhydrin assay, at 30°C in a total volume of 5 mL. The reaction was started by addition of 0.5 mL of PDF in dilution solution to 4.5 mL of solution containing 5.5 mM *N*-formyl-Met-Ala-OH as the substrate. The, 800 µL aliquots were withdrawn from the reaction mixture every 5 min and added to 400 µL of DMBA reagent. After 20 min at room temperature, the absorbance was measured at 415 nm.

Site-directed mutagenesis

For the mutagenesis work, the pBAD-PDF_{Etag} construct was used as a template. A single mutant (E142A) and a double mutant (C138A/H141A) were constructed using

the QuikChange Site-Directed Mutagenesis Kit of Stratagene, following the protocol provided by the manufacturer. E142 was mutated to A with primer 5'-GCCATATGTATT-CAGCAT**GCG**ATGGATCACCTGG-3' and its complementary primer. The double mutant was prepared by mutating C138 to A and H141 to A with the primer 5'-GGTCTGTTAGCGATC**GCG**ATTCAG**GCG**GAGATGGATCACC-3' and its complementary primer. The mutated nucleotides are indicated in bold. In order to check the efficacy of the mutagenesis, a silent mutation was created in each primer to introduce a restriction site. The recognition sites of *Pvu*I (double mutant primer) and *Nde*I (single mutant primer) are underlined. The resulting plasmids were transformed in *E. coli* MC1061 and verified by SDS polyacrylamide gel electrophoresis, sequencing and activity assays.

References

1. **Anonymous Editorial.** 2009. Closing in on catalysis. Nat Chem Biol **5**:515.
2. **Andersen, K. B. and K. von Meyenburg**. 1980. Are growth rates of *Escherichia coli* in batch cultures limited by respiration? J. Bacteriol. **144**:114-123.
3. **Becker, S., H. U. Schmoldt, T. M. Adams, S. Wilhelm, and H. Kolmar**. 2004. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. Curr. Opin. Biotechnol. **15**:323-329.
4. **Bingel-Erlenmeyer, R., R. Kohler, G. Kramer, A. Sandikci, S. Antolic, T. Maier, C. Schaffitzel, B. Wiedmann, B. Bukau, and N. Ban**. 2008. A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature **452**:108-111.
5. **Boersma, Y. L., M. J. Droge, and W. J. Quax**. 2007. Selection strategies for improved biocatalysts. FEBS journal **274**:2181-2195.
6. **Bornscheuer, U. T., J. Altenbuchner, and H. H. Meyer**. 1999. Directed evolution of an esterase: screening of enzyme libraries based on pH-Indicators and a growth assay. Bioorg. Med. Chem. **7**:2169-2173.
7. **Bornscheuer, U. T. and M. Pohl**. 2001. Improved biocatalysts by directed evolution and rational protein design. Curr. Opin. Biotechnol. **5**:137-143.
8. **Brakmann, S. and A. Schwienhorst**. 2004. Evolutionary methods in biotechnology: clever tricks for directed evolution: Wiley-VCH Verlag GmbH,Weinheim.
9. **Denault, M. and J. N. Pelletier**. 2007. Protein Library Design and Screening: Working Out the Probabilities, in Protein Engineering Protocols, vol 352, p. 127-154: Humana Press.
10. **Doi, N. and H. Yanagawa**. 2001. Genotype-phenotype linkage for directed evolution and screening of combinatorial protein libraries. Com. Chem. High T. SCR **4**:497-509.
11. **Forrer, P., S. Jung, and A. Plnckthun**. 1999. Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. Curr. Opin. Struc. Biol. **9**:514-520.

12. **Goddard, J. P. and J. L. Reymond**. 2004. Recent advances in enzyme assays. Trends Biotechnol. **22**:363-370.

13. **Griffiths, A. D. and D. S. Tawfik**. 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. EMBO J. **22**:24-35.

14. **Guzman, F., S. Barberis, and A. Illanes**. 2007. Peptide synthesis: chemical or enzymatic. Electronic J. Biotech. **10**:279-314.

15. **Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith**. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. J. Bacteriol. **177**:4121-4130.

16. **Hempfling, W. P. and S. E. Mainzer**. 1975. Effects of varying the carbon source limiting growth on yield and maintenance characteristics of Escherichia coli in continuous culture. J. Bacteriol. **123**:1076-1087.

17. **Hirsh, J. and R. Schleif**. 1973. In vivo experiments on the mechanism of action of -arabinose C gene activator and lactose repressor. J. Mol. Biol. **80**:433-444.

18. **Hodgson, D.** 1982. Glucose repression of carbon source uptake and metabolism in *Streptomyces coelicolor* A3(2) and its perturbation in mutants resistant to 2-deoxyglucose. J. Gen. Microbiol. **128**:2417-2430.

19. **Kazlauskas, R. J. and U. T. Bornscheuer**. 2009. Finding better protein engineering strategies. Nat. Chem. Biol. **5**:526-529.

20. **Lazennec, C. and T. Meinnel**. 1997. Formate dehydrogenase-coupled spectrophotometric assay of peptide deformylase. Anal. Biochem. **244**:180-182.

21. **Leemhuis, H., R. M. Kelly, and L. Dijkhuizen**. 2009. Directed evolution of enzymes: Library screening strategies. IUBMB Life **61**:222-228.

22. **Lin, H. and V. Cornish**. 2002. Screening and selection methods for large-scale analysis of protein function. Angew. Chem. Int. Ed. Engl. **41**:4402-4425.

23. **Liu, M., T. Durfee, J. E. Cabrera, K. Zhao, D. J. Jin, and F. R. Blattner**. 2005. Global transcriptional programs reveal a carbon surce foraging strategy by *Escherichia coli*. J. Biol. Chem. **280**:15921-15927.

24. **Mazel, D., S. Pochets, and P. Marliere**. 1994. Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO **13**:914-923.

25. **Meinnel, T., C. Lazennec, and S. Blanquet**. 1995. Mapping of the active site zinc ligands of peptide deformylase. J. Mol. Biol. **254**:175-183.

26. **Moris-Varas, F., A. Shah, J. Aikens, N. P. Nadkarni, J. D. Rozzell, and D. C. Demirjian**. 1999. Visualization of enzyme-catalyzed reactions using pH indicators: rapid screening of hydrolase libraries and estimation of the enantioselectivity. Bioorg. Med. Chem. **7**:2183-2188.

27. **Olsen, M., B. Iverson, and G. Georgiou**. 2000. High-throughput screening of enzyme libraries. Curr. Opin. Biotechnol. **11**:331-337.

28. **Paliy, O. and T. Gunasekera**. 2007. Growth of *E. coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents. Appl. Microbiol. Biotechnol. **73**:1169-1172.

29. **Pohn, B., J. Gerlach, M. Scheideler, H. Katz, M. Uray, H. Bischof, I. Klimant, and H. Schwab**. 2007. Micro-colony array based high throughput platform for enzyme library screening. J. Biotech. **129**:162-170.

30. **Ragusa, S., P. Mouchet, C. Lazennec, V. Dive, and T. Meinnel**. 1999. Substrate recognition and selectivity of peptide deformylase. similarities and differences with metzincins and thermolysin. J. Mol. Biol. **289**:1445-1457.

31. **Reisinger, C., F. van Assema, M. Schnrmann, Z. Hussain, P. Remler, and H. Schwab**. 2006. A versatile colony assay based on NADH fluorescence. J. Mol. Catal. B: Enzymatic **39**:149-155.

32. **Roberts, R. W.** 1999. Totally in vitro protein selection using mRNA-protein fusions and ribosome display. Curr. Opin. Biotechnol. **3**:268-273.

33. **Schleif, R.** 2000. Regulation of the arabinose operon of *Escherichia coli*. Trends Genet. **16**:559-565.

34. **Sonke, T. and et al.** 2009. Enzyme assays: high-throughput screening, genetic selection and fingerprinting. Wiley-VCH.

35. **Stoscheck, C.** 1990. Quantitation of protein. Method. Enzymol. **182**:50-69.

36. **Sylvestre, J., H. Chautard, F. Cedrone, and M. Delcourt**. 2006. Directed evolution of biocatalysts. Org. Process Res. Dev. **10**:562-571.

37. **Tao, H., C. Bausch, C. Richmond, F. R. Blattner, and T. Conway**. 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. **181**:6425-6440.

38. **Taylor, S., P. Kast, and D. Hilvert**. 2001. Investigating and engineering enzymes by genetic selection. Angew. Chem. Int. Ed. Engl. **40**:3310-3335.

39. **Turner, N. J.** 2003. Directed evolution of enzymes for applied biocatalysis. Trends Biotechnol. **21**:474-478.

40. **Turner, N. J.** 2009. Directed evolution drives the next generation of biocatalysts. Nat. Chem. Biol. **5**:567-573.

41. **Wahler, D. and J. L. Reymond**. 2001. Novel methods for biocatalyst screening. Curr. Opin. Biotechnol. **5**:152-158.

42. **Xu, Z., W. Shen, H. Chen, and P. Chen**. 2005. Effects of medium composition on the production of plasmid DNA vector potentially for human gene therapy. Journal of Zhejiang University science **6B**:396-400.

Chapter 5

Saturated mutagenesis of EcPDF to broaden its substrate range

Abstract

The application of peptide deformylase in biocatalysis is restricted by its narrow substrate specificity. In this Chapter, a site-saturation mutagenesis approach was followed to broaden the substrate range of the $EcPDF_{Etag}$, which is a modified version of *E. coli* peptide deformylase. The changes were directed toward four non-conserved residues known to be involved in substrate binding. The library was obtained by combinatorial randomization of these residues. The application of a growth selection method (described in Chapter 4) using *N*-formyl-Phe-Leu-OH as substrate yielded a mutant with 2.5-fold increased activity. Following the same approach, the activity of $EcPDF_{Etag}$ towards *N*-formyl-Leu-Phe-NH_2 could be increased by two-fold. Sequencing of the mutants revealed the presence in both cases of a single mutation at the position of Cys-138, revealing the importance of this residue in substrate recognition. Subsequently, the activities of 20 different *E. coli* transformants expressing $EcPDF_{Etag}$ variants, covering all possible proteinogenic amino acids at position 138, was determined and five new substitutions at this position leading to active mutants were discovered. These clones displayed new substrate specificities and were partially characterized.

Introduction

Over the last decades, peptide deformylase (PDF) has been recognized as an interesting target for the development of antibacterial and anticancer drugs. Several crystal structures of PDFs of bacterial origin were solved with the aim to assist the design of new inhibitors for the enzyme, and a nmber of publications on this topic have appeared [9;17 15 2 21 8]. Data on the use of PDF for other applications are very limited. Only one publication describes the use of *E. coli* PDF for the resolution of racemates of N-formylated α- and β-amino acids, α-amino acid amides and α-aminonitriles, which are interesting building blocks for organic synthesis [23]. An important observation was that PDF has good regio- and enantioselectivity [19;23].

Substrate recognition and selectivity of PDF from *E. coli* were comprehensively studied by Ragusa and coworkers [20], who identified a set of enzyme-substrate interactions which are responsible for its tight specificity. The presence of a hydrophobic pocket at the S_1' site for methionine side-chain binding allows efficient deformylation of

peptides with a formylated methionine only. *N*-formylated peptides with phenylalanine, leucine, tyrosine and isoleucine in position P_1' are accepted but with a catalytic efficiency that is nearly two orders of magnitude lower (Table 1) [20]. Consequently, for developing new applications, PDF needs to be changed from a highly specific enzyme into an enzyme with relaxed substrate specificity that will effectively deformylate *N*-formylated peptides with other P_1'side chains than methionine.

The hydrophobic S_1' pocket of *E. coli* PDF (Figure 1) is built of three subparts [20]: 1) a ceiling, framed by the side chains of residues Glu-88 and His-132; b) the floor, characterized by Ile-44 and Cys-129; c) the back, lined by the side chains of Ile-86, Leu-125 and Ile-128.

Table 1. Effect of the nature of the first amino acid residue (P_1')on the hydrolysis of a combinatorial peptide library by PDF from *E. coli* [20]

Formylated N-terminal amino acid	k_{cat}/K_m ($M^{-1}s^{-1}x10^3$)	Relative k_{cat}/K_m
Methionine	131±11	100
Phenylalanine	37±4	28
Leucine	3.0±0.3	2.3
Tyrosine	2.4±0.3	1.8
Isoleucine	1.8±0.08	1.4
Glutamine	1.8 ±0.02	1.4
Alanine	0.84±0.08	0.6
Valine	0.19±0.02	0.14
Histidine	0.07±0.01	0.05
Lysine	0.05±0.01	0.04
Serine	0.04±0.01	0.03
Threonine	0.034±0.004	0.026
Glutamic acid	0.018±0.002	0.014
Glycine	0.017±0.002	0.013
Arginine	0.015±0.002	0.011
Asparagine	0.011±0.001	0.008
Aspartic acid	0.0009±0.00001	0.0007
Proline	«0.00003	«0.00002
Tryptophan	NS	NS

NS, no substrate

Only two of these residues (Glu-88 and His-132) are involved in the hydrogen bonding network that is crucial for correct three-dimensional folding and for binding of the metal ion that is essential for catalysis and these are highly conserved among PDFs of different origins. The other residues that shape the S_1' pocket are not conserved

although the selectivity of different peptide deformylase is identical [14;16](see Chapter 3). Therefore, the high specificity is not primarily due to specific interactions with side chains of single amino acids but

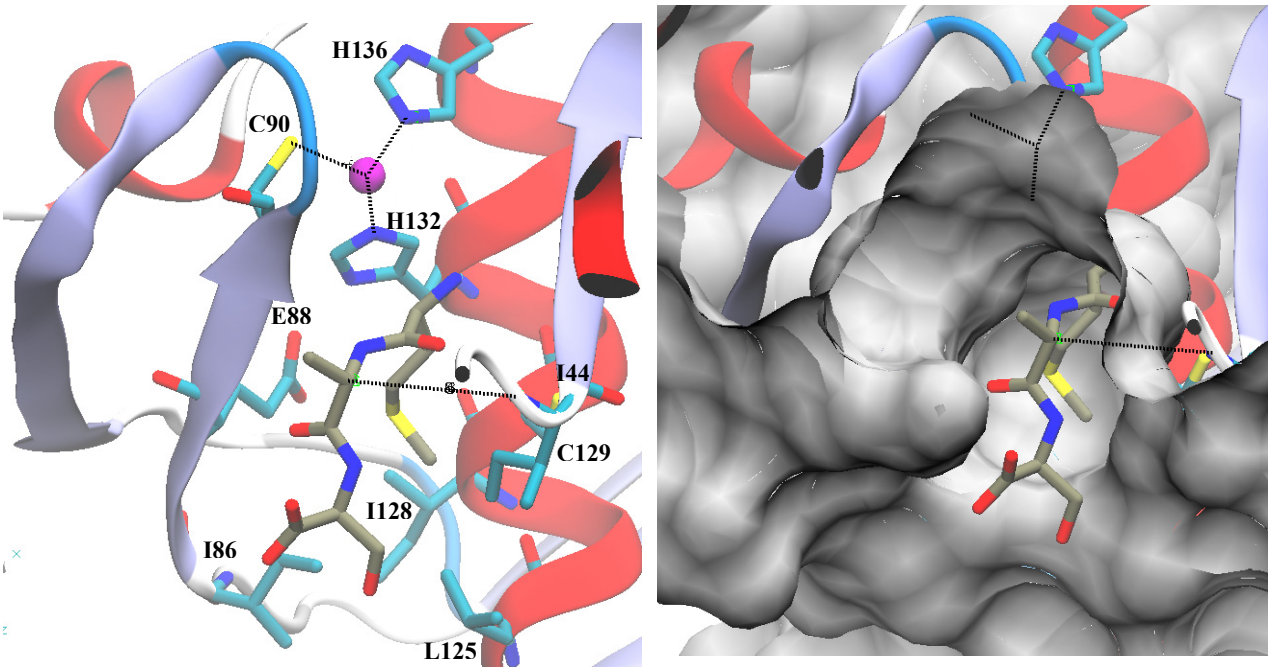


Figure 1. Representations of the active site of $EcPDF_{wt}$ with H-Met-Ala-Ser-OH as the ligand in the active site.. (A) Residues forming the hydrophobic pocket are shown as sticks. Ni^{2+} is coordinated by three conserved residues (His-132, His-136, Cys-90). (B) Structure of the pocket shown in surface display mode. The figures were generated with VMD software using pdb file 1BS6.

should be attributed to the overall size and hydrophobicity of the crevice. Based on this assumption and on published work on substrate recognition and docking experiments, as well as on the design of inhibitors for peptide deformylase [6;10;14;15;20;22], we decided to target the amino acids Ile-86, Leu-125, Ile-128 and Cys-129 for expanding the substrate range.

Different tools are available to alter enzyme specificity and to create enzyme variants that are able to accept a larger range of substrates [18]. Site-directed mutagenesis has successfully been used for redesigning substrate specificity in different classes of enzymes [11-13;24]. This approach is based on detailed knowledge of the enzyme structure and mechanism [18]. Use of a more random approach followed by different rounds of selection is an alternative approach. In this case, knowledge about the structure of the enzyme and interactions in the active site is not required, and mutations that influence the substrate specificity can be obtained in regions not directly

in contact with the substrate. A drawback of this technique is the need to establish a selection or screening method that is suitable for evaluating the large numbers of mutants that are typically needed for finding the desired variants in random directed evolution libraries[1].

In this Chapter, we report the use of a hybrid approach for the development of PDF variants with altered substrate specificity. In this method, four of the residues forming the hydrophobic S_1' pocket in *E. coli* PDF, i.e. Ile-95, Leu-134, Ile-137 and Cys-138, were simultaneously randomized by saturated mutagenesis (Chapter 2). These correspond to Ile-86, Leu-125, Ile-128 and Cys-129 in the wild-type form of *E. coli* PDF. The choice for simultaneous randomization was supported by the synergistic effects that could possibly occur and that can not be reliably predicted for mutations applied to individual positions [3;5;25]. Furthermore, the mutagenesis only targets positions which on the basis of the structure are expected to influence the selectivity of the enzyme. A gene library comprising about 220,000 mutants was generated and clones exhibiting the desired phenotype were identified by a growth selection assay. Analysis of the positive clones revealed a dominant role of the residue at position Cys-138 in determining the substrate specificity of PDF.

Materials and methods

Microorganisms, plasmids and cultivation conditions

Three different *Escherichia coli* strains were used as host for the propagation and expression of plasmid pBAD-PDF_{Etag} and its mutants. Plasmid-containing *E. coli* ElectroMAX DH5α-E™ cells (Invitrogen), *E. coli* TOP10 (Invitrogen) and *E. coli* MC1061 [*hsdR2 hsdM+ hsdS+ araD139 Δ(ara-leu)7697 Δ(lac)X74 galE15 galK16 rpsL (StrR) mcrA mcrB1*] were cultivated at 28, 30 or 37°C in LB medium supplemented with 100 mg/L carbenicillin (LB-carb) for 16-18 h. The expression of the *def* gene, which encodes PDF, and its mutants, was induced in TOP10 and MC1061 cells by addition of 0.002 % (w/v) arabinose when the OD_{620} of the culture reached 0.5-1.0.

During the selection assays, strain *E. coli* MC1061 was cultivated on minimal medium plates (pH 7.3) that contained per liter: 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 240 mg $MgSO_4$, 11 mg $CaCl_2$, 50 mg L-alanine, 50 mg L-arginine, 500 mg L-aspartic acid, 500 mg L-glutamic acid, 50 mg glycine, 50 mg L-histidine, 50 mg L-isoleucine, 50 mg L-lysine, 50 mg L-methionine, 50 mg L-phenylalanine, 50 mg L-

proline, 50 mg L-serine, 50 mg L-threonine, 20 mg L-tryptophan, 20 mg L-tyrosine, 50 mg L-valine, 2 g glucose, 40 mg thiamine, 100 mg of carbenicillin, 5 mL of a trace elements solution and 20 g agarose. The trace elements solution contained per liter 780 mg $Ca(NO_3)_2 \cdot 4H_2O$, 200 mg $FeSO_4 \cdot 7H_2O$, 10 mg $ZnSO_4 \cdot 7H_2O$, 10 mg H_3BO_3, 10 mg $CoCl_2 \cdot 6H_2O$, 10 mg $CuSO_4 \cdot 5H_2O$, 4 mg $MnSO_4 \cdot H_2O$, 3 mg $Na_2MoO_4 \cdot 2H_2O$ and 2 mg $NiCl_2 \cdot 6H_2O$. For control tests 50 mg/L L-leucine was added to the medium. The L-arabinose concentration for PDF expression in growth selection experiments varied from 0.002 to 0.2 % (w/v).

Electroporation of *E. coli* MC1061 cells

E. coli MC1061 cells were made electrocompetent by inoculating a single colony into 5 mL LB medium, followed by overnight growth at 28°C. Next, 2.5 mL of this culture was transferred to 500 mL LB medium and grown at 37°C to an OD_{620} of about 0.5. Cells were then harvested by centrifugation for 20 min at 4,400x *g* in Beckman JLA 8.1000 and kept at 4°C during subsequent steps. The pellet was washed three times in ice-cold water suspended in volume of ice-cold water equal to the pellet volume. This yielded electrocompetent cells that were stored at -20°C.

For electroporation, 0.2 ng of plasmid DNA was added to 50 µL of electrocompetent cells followed by application of a pulse in an electroporation cuvette (400 ohms, 2.5 kV, 25 µF). After addition of 1 mL of SOC medium (containing per liter: 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 2.5 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) the cells were incubated at 37°C for one h before plating appropriate dilutions on LB plates supplemented with 100 mg/L carbenicillin.

Construction of a four site saturated expression library

Forward and reverse primers with random codons (NNN) at the target positions were designed to recombine and randomize the codons for each of the four amino acid residues that were expected to influence the substrate specificity of PDF (Ile-95, Leu-134, Ile-137 and Cys-138) (Table 2).

Two separate PCR reactions (50 µL each) were performed for each of the four codons using Herculase II Fusion Enzyme (0.5 µl, Stratagene) and accompanying buffer, 30 ng pBAD-PDF_{Etag} template, 1 mM dNTPs and either a codon-specific saturated (NNN) forward primer in combination with the reverse external primer or a

codon-specific saturated (NNN) reverse primer in combination with the forward external primer (all primers: 250 nM)(Figure 2). All 8 reactions (4 residues x 2 reactions per

Table 2. Synthetic oligonucleotides used for site saturated mutagenesis. Boldface – codon to be saturated.

Primer	Codon	Nucleotide sequence (5' → 3')
PDFMUT 1for	ATT	GCGGCGAAACAGGC**NNN**GAAGAAGGTTGCC
PDFMUT 2for	CTG	GAACTGGAAGCAGACGGT**NNN**TTAGCCATC
PDFMUT 3for	ATC	GACGGTCTGTTAGCC**NNN**TGTATTCAGC
PDFMUT 4for	TGT	CGGTCTGTTAGCCATC**NNN**ATTCAGCATGAG
PDFMUT 1rev	AAT	GCAACCTTCTTC**NNN**GCCTGTTTCGCCG
PDFMUT 2rev	CAG	GATGCCTAA**NNN**ACCGTCTGCTTCCAGTTC
PDFMUT 3rev	GAT	CATGCTGAATACA**NNN**GGCTAACAGACCGTCT
PDFMUT 4rev	ACA	TCTCATGCTGAAT**NNN**GATGGCTAACAGACCG
PDFex 1for		TCGCAACTCTCTACTGTTTCTCCATACCC
PDFex1rev		CCTGGCAGGTTCCCTACTCTCGCATGG
att-Primer forward		AGGAGGAATTAACCATGTCCGTGCTTCAAGTGTTACATATTCC
att-Primer reverse		TTAAGCCCGGGCTTTCAGACGATCCAGTTTTTC

residue) were started with an initial denaturation for 2 min at 95°C, followed by 30 cycles of 20 s at 95°C, 20 s at 55°C, and 30 s at 72°C, and ended with an additional incubation of 3 min at 72°C. After checking with agarose gel electrophoresis whether a single product was formed, the 8 amplification products were purified using the QIAquick PCR Purification Kit (Qiagen) and the final concentration of each single fragment was determined with the Qubit fluorometer in combination with Quant-iT dsDNA HS Assay Kit (Invitrogen). After this, an equimolar mixture of all fragments was added to a solution containing Herculase II Fusion Enzyme (0.5 µL), 1 mM dNTPs and accompanying buffer (50 µL). No additional primers or DNA template were added in this step. After an initial denaturation for 2 min at 95°C, the fusion PCR was performed by 30 cycles of 20 s at 95°C, 20 s at 55°C, and 30 s at 72°C, followed by a final extension of 3 min at 72°C. The product of this fusion PCR was purified via preparative gel electrophoresis followed by extraction of the fragment from the gel with the QIAquick Gel Extraction Kit (Qiagen). Subsequently, the product was amplified (200 ng) in a new PCR reaction (50 µL) containing Herculase II Fusion Enzyme (0.5 µl) and accompanying buffer, 1 mM dNTPs, 250 nM att-Primer forward and 250 nM att-Primer

reverse, using the same amplification conditions as above in order to add the recombination sites needed for cloning. The resulting product of 605 nt was then purified using the Qiaquick PCR Purification Kit (Qiagen) and cloned into pDONR/Zeo according to standard Gateway technology protocols. The resulting plasmid collection was transformed into *E. coli* ElectroMAX DH5α-E by electroporation. Transformants were selected by plating on LB plates containing zeomycin with incubation at 28°C for 16-18 h, resulting in a library of about 400,000 clones. All colonies from this entry library were collected in LB liquid medium, after which plasmid DNA was isolated from 150 µL aliquots of the cell suspension using QiaPrep Spin Miniprep kit (Qiagen). The resulting plasmid preparations were stored and used as the mutant library.

After checking the quality of this entry library by sequencing of 14 randomly picked clones, a standard LR recombination reaction (a recombination reaction between *attL* and *attR* sites in standard Gateway cloning vectors) was applied to transfer the gene library from pDONR/Zeo to pBAD/*Myc*-His-DEST (see Materials and Methods, Chapter 2) with a molar ratio of destination vector to entry vector of 1:2 (150:300 ng). After electroporation of *E. coli* MC1061, transformants were selected by plating on LB medium containing 100 µg/ml carbenicillin, followed by overnight incubation at 28°C.

Selection of active mutants and screening for increased activity

Active mutants were selected for their ability to allow the host (*E. coli* MC1061, leucine auxotrophic strain) to grow at 37°C on 1 mM *N*-formyl-Leu-Phe-NH_2, *N*-formyl-Phe-Leu-OH or *N*-formyl-Asp-Leu-OH as the sole source of leucine using a minimal solid medium supplemented with L-arabinose (0.02 or 0.2 % w/v) and glucose (0.2% w/v). After 5 days of growth, single colonies were transferred from the agar plates to 96-well microtiter plates (MTPs) filled with 200 µL LB-carb medium per well. After overnight incubation at 28°C with 200 rpm orbital shaking, 100 µL samples of the cultures were transferred to new 96-well plates followed by addition of glycerol (8% w/v) and storage at -80°C as master plate. Cells of the remaining part of the cultures were collected by centrifugation, resuspended in 100 µL physiological salt solution (0.91% w/v of NaCl) containing 8% glycerol and stored at -80°C. For re-testing, a sample from the master plate was thawed and plated on minimal medium under the same conditions as applied in the first selection, and cells were grown at 37°C for 2 days.

Plasmids were isolated from growing colonies by the QiaPrep Spin Miniprep kit (Qiagen) and subsequently reintroduced into *E. coli* MC1061 according to the protocols described above.

Positive hits, obtained via selection for auxotrophic complementation on minimal medium, were further investigated by determination of the PDF activity. In order to do so, the colonies growing on minimal medium were transferred to 30 mL LB-carb. Expression of the $PDF_{E\text{-}tag}$ variant was induced by adding 0.002% L-arabinose when the OD_{620} reached 0.6. Cells were harvested after overnight incubation at 28°C by centrifugation at 5,000 x g for 15 min and 4°C. To prepare a cell-free extract (CFE), cells were resuspended in 1 mL lysis buffer composed of 50 mM MOPS pH 7.5, 0.1 mg/mL DNAse I grade II (Roche), 2 mg/mL lysozyme (Bachem), 1.23 mg/mL $MgSO_4 \cdot 7H_2O$, 10 µg/mL catalase from bovine liver (Sigma) and 1 mM TCEP and stored overnight at -20°C. Then the cell suspension was thawed, incubated for one hour at room temperature and subsequently CFE was obtained by centrifugation at 33,300 x *g* for 1 h at 4°C. Expression levels of the PDF mutant enzymes was checked by SDS-PAGE as described below.

Deformylase activity was measured at 30°C in a total volume of 500 µL with 50 µL of different dilutions of PDF (CFE or whole cells) as described under the paragraph Activity assays. The incubation time was varied based on the expected activity of the mutant clones on the different substrates. Analysis of substrate and product was performed using the HPLC method described below.

Single-site saturation mutagenesis of Cys-138

Construction of a site-saturation library of amino acid Cys-138 in $EcPDF_{E\text{-}tag}$ was performed by Sloning Biotechnology. The 17 variants (two mutants were available already, see Results) were cloned into pBAD/*Myc*-His-DEST. Transformation of *E. coli* TOP10 and plating on LB-carb plates, followed by overnight incubation at 28°C, yielded the desired transformants.

Protein electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was routinely performed in NuPAGE 4 to 12% Bis-Tris gels (Invitrogen) under reducing

conditions with MOPS-SDS running buffer. The gels were stained with the SimplyBlue SafeStain (Invitrogen).

Enzyme purification

E. coli TOP10 cells carrying plasmid pBAD-PDF_{Etag} and its variants were cultivated at 28-30°C in LB medium supplemented with 100 mg/L carbenicillin for 16-18 h. The expression of the *def* gene and its variants was induced when the OD_{620} reached 0.5-1.0 by the addition of 0.002 % arabinose. Cells were harvested by centrifugation and about 2 g (wet weight) of cell paste was suspended in 10 ml buffer (20 mM HEPES/KOH, 100 mM KCl, pH 7.7 supplemented with 10 µg/mL catalase from bovine liver). After disintegration by sonication (Branson B12, 10 min) at 0°C, cell-free extract was obtained by centrifugation at 33,300 x *g* for 1 h at 4°C. CFE was subsequently used for enzyme purification.

Purification of all PDF mutants was executed using an ÄKTA explorer system (GE Healthcare) equipped with a Mono Q HR 5/5 column (GE Healthcare). The procedure is described in Chapter 2.

Activity assays

The activity of the PDF variants was assayed towards several substrates in 450 µL reaction mixtures containing 110 mM MOPS/NaOH buffer (pH 7.2), 300 mM NaCl, 0.1 mg/mL bovine liver catalase and 0.6 mM of the desired substrate. To this substrate solution was added 50 µL of an appropriate dilution of PDF in 20 mM MOPS/NaOH, pH 7.7, 100 mM NaCl, 10 µg/mL bovine liver catalase, and 2 mM TCEP. After starting the reaction by the addition of the PDF, 100 µL aliquots were withdrawn from the reaction mixture and added to 100 µL of 1 M phosphate buffer (H_3PO_4-NaOH), pH 2.66, to stop the reaction. Sampling times were adjusted in accordance with the activity of the PDFs for the substrate ensuring activity measurements in the initial linear part of the progress curve. The substrates that were used for the activity tests were: *N*-formyl-Leu-Phe-NH_2; *N*-formyl-Phe-Leu-OH; *N*-formyl-Val-Val-OH; *N*-formyl-Tyr-Leu-Phe-NH_2; *N*-formyl-Met-Ala-OH; *N*-formyl-Lys-Phe-NH_2; *N*-formyl-Phe-Phe-NH_2 provided by Chiralix (Nijmegen, the Netherlands). Samples were analyzed using the HPLC method described below.

HPLC analyses

HPLC analysis was carried out using a stainless-steel analytical column (50 mm length, 4.6 mm ID) packed with Inertsil ODS-3 material, 3 µm particle size (Alltech Applied Science). The flow was 2 mL/min. Detection was performed at 40°C at a wavelength of 210 nm. For HPLC detection of *N*-formyl-Met-Ala-OH (2.21 min), H-Met-Ala-OH (0.63 min), *N*-formyl-Val-Val-OH (1.86 min), H-Val-Val-OH (0.51 min), *N*-formy-Tyr-Leu-Phe-NH_2 (2.53 min), the injection volume was 10 µL and the following gradient of acetonitrile in 10 mM H_3PO_4 was used: t=0-0.5 min, 0.5 % (v/v) acetonitrile isocratic; t=0.5-3 min, 0.5%-50% acetonitrile linear increase; t=3-3.1 min, 50%-0.5% acetonitrile linear decrease. For HPLC analysis of *N*-formyl-Phe-Leu-OH (2.59 min), H-Phe-Leu-OH (1.56 min), *N*-formyl-Lys-Phe-NH_2 (0.91 min), H-Lys-Phe-NH_2, N-formyl-Leu-Phe-NH_2 (2.34 min), H-Leu-Phe-NH_2 (1.26 min), *N*-formyl-Asp-Leu-OH (1.73 min), H-Asp-Leu-OH (0.8 min) the injection volume was 5 µL and the following gradient of acetonitrile in 10 mM H_3PO_4 was used: t=0-3 min, 5-65% (v/v) acetonitrile linear increase; t= 3-3.1 min, 65-5% acetonitrile linear decrease.

Modeling of C138A and C138V mutants

In order to evaluate the effects of the mutations C138A and C138V on the active site architecture, their amino acid sequences were used to build three-dimensional models with the SWISS-MODEL server (http://swissmodel.expasy.org//SWISS-MODEL.html).

Results and discussion

Structure-inspired saturated mutagenesis

In order to change the substrate specificity of the $EcPDF_{Etag}$, residues potentially involved in substrate recognition were identified by inspecting PDF crystal structures and the reaction mechanism. The substrate binding site of PDFs is formed by 7 residues [14;20], and of these, four residues were chosen as targets for the structure-guided saturated mutagenesis. The importance of residues His-132 and Glu-88, which together delineate the ceiling of the pocket in EcPDF, was demonstrated by Meinnel and coworkers [6]. These residues are strictly conserved in PDFs and are involved in the correct folding of the 3D structure of the enzyme via a complex hydrogen bonding network. His-132 also serves as a ligand for the metal ion (Figure 1) and Glu-88 is

involved in a series of hydrogen bonds which reinforce the enzyme structure in the vicinity of the metal ion. Therefore they were omitted from the set of target positions for mutagenesis. Another residue that is part of the S1' pocket is Ile-53, which appears to be a hydrogen bond donor to the second carbonyl of the substrate and most likely plays no important role in the recognition of the P_1' side chain. Therefore, the sites that were selected as mutagenesis targets in the PDF_{Etag} were Ile-95, Leu-134, Ile-137 and Cys-138.

An interesting feature of these residues is that they are not conserved, although the substrate specificity of peptide deformylases originating from different organisms is strictly identical [20] and Chapter 3. For this reason it was decided to simultaneously mutate these four sites, creating a mutant library of high sequence diversity. The advantage of mutating in a combinatorial way instead of preparing single mutants is reported in various papers showing that mutants having multiple amino acid substitutions have larger differences in substrate specificity compared to single mutants. The effects of individual mutations on substrate selectivity can be influenced by the identity of other residues that are near to the targeted region due to combined effects on binding site geometry and substrate orientation [1;25]. In fact, we expected that the substrate specificity of PDF will be influenced by the overall size and hydrophobic character of the crevice more than by the identity of single amino acid side chains. Only hydrophobic interactions are responsible for formation of the hydrophobic pocket that nicely accommodates the methionine side chain (His-141, Cys-138, Ile-137, Leu-134, Ile-95, Glu-97, Ile-53). Hydrogen bonds are formed only between atoms of the peptide backbone of the substrate and Gly-53, Ile-53, Gly-98 and Arg-106. This is consistent with the low sequence specificity of the enzyme for the P_2' and further amino acid residues of the nascent polypeptide chain.

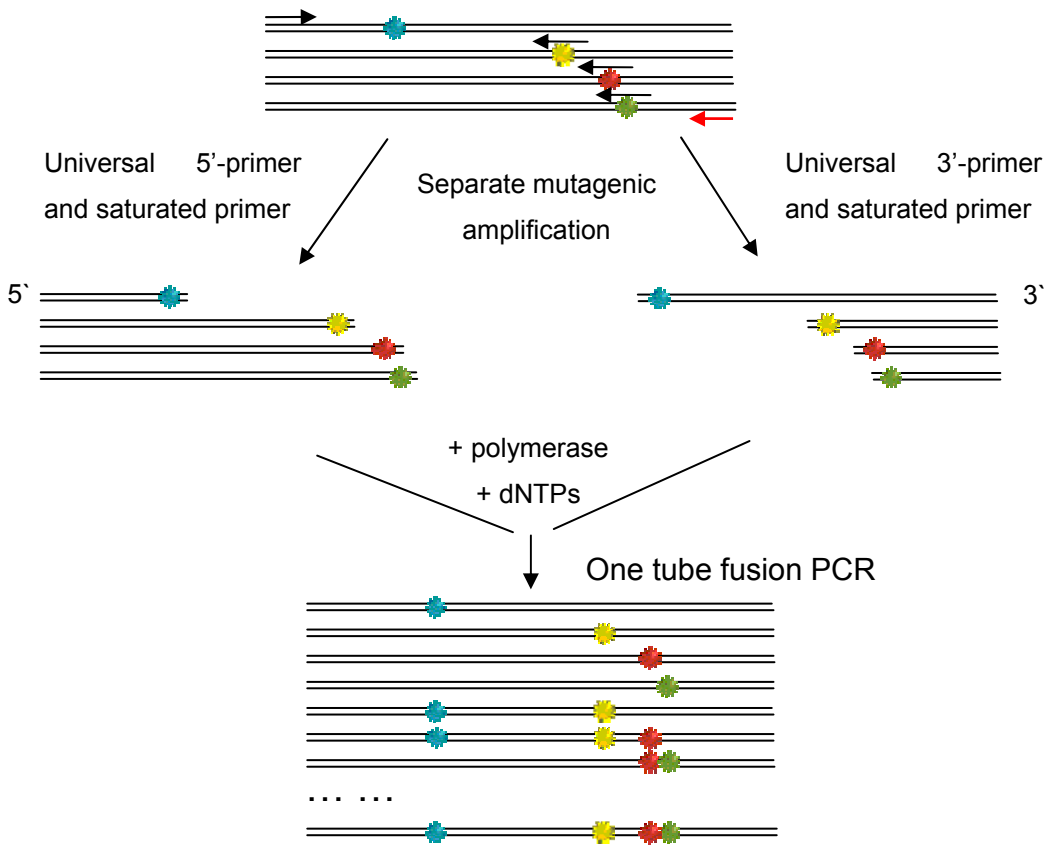


Figure 2. Schematic representation of the mutagenesis procedure used in this study. In the first step two separate PCR reactions were performed for each of the four codons using a random codon in the target position of the mutagenic primer and an universal external primer. After purification of the amplicons, an equimolar mixture of all fragments was added to a solution containing polymerase. The resulting product was then purified, multiplied with recombination sites containing primer, and cloned into pDONR/Zeo.

The mutant library was constructed with the Saturated Mutation Primer PCR (SMPP) Recombination method [4] applying the $EcPDF_{Etag}$ gene as template and using saturated mutagenic forward and reverse primers for simultaneous randomization of the four target amino acid residues. This enabled the exploration of all possible combinations and permutations of interesting amino acid residues at the 4 target sites, thereby including possible synergetic interactions which may be caused by compensating geometric effects that may be required for fitting different P1' side chains. Cloning of the obtained fragments into vector pBAD/*Myc*-His-DEST via in vitro recombination resulted in a mutant library of about 400,000 colonies. The quality of the library was checked via sequencing of the PDF genes of 42 randomly picked colonies. In total 11 of these genes revealed the wild-type sequence, 22 carried a single mutation, 7 carried two mutations

and 2 were triple mutants. Despite the presence of wild-type PDF genes, the library was sufficiently rich in molecular diversity. Furthermore the use of a growth selection method for the identification of the best mutants allows an extremely high-throughput with the exclusive survival of desired variants and removal of wild-type enzymes before further analysis.

Selection of mutants with altered substrate specificity

The development of a selection method for PDF variants with increased activity toward a formylated dipeptide containing leucine is described in Chapter 4. This approach is based on the complementation of *E. coli* MC1061 leucine auxotrophy. For this purpose, cells are grown on a medium containing leucine in a form of a formylated dipeptide. The successful utilization of this substrate can be achieved only if it is unmasked to the dipeptide by a peptide deformylase because this is required for subsequent hydrolysis into the free amino acids by cellular peptidases. The substrates used in the selection procedure were *N*-formyl-Phe-Leu-OH, *N*-formyl-Leu-Phe-NH_2 and *N*-formyl-Asp-Leu-OH (Figure 3). These substrates were chosen because a study by Ragusa et al. [20] showed that the catalytic efficiencies [k_{cat}/K_M] of $EcPDF_{wt}$ for these substrates differ by more than 4 orders of magnitude [20].

The first selection was conducted on agar plates containing *N*-formyl-Phe-Leu-OH as the only source of leucine and 0.2% L-arabinose to induce protein expression. A total of about 400,000 colony forming units (CFUs) were plated on this. After 5 days at 37°C, 15 of these actually formed colonies. Re-streaking of these positive colonies showed that 10 of them were still able to grow on the selective plates, while the remaining 5 were false positives. Subsequently, plasmid borne activity was proven for 3 of these 10 hits by isolation of the plasmid from these primary hits followed by reintroduction into *E. coli* MC1061 and a second selection round on the same medium. This procedure allowed us to discriminate among the clones that were able to grow due to a mutation carried on the plasmid and the ones favouring growth due to chromosomal mutations that might theoretically appear due to such diverse effects as better uptake of the substrate, increased PDF overexpression, or higher peptidase activity leading to more leucine formation from formylated dipeptides.

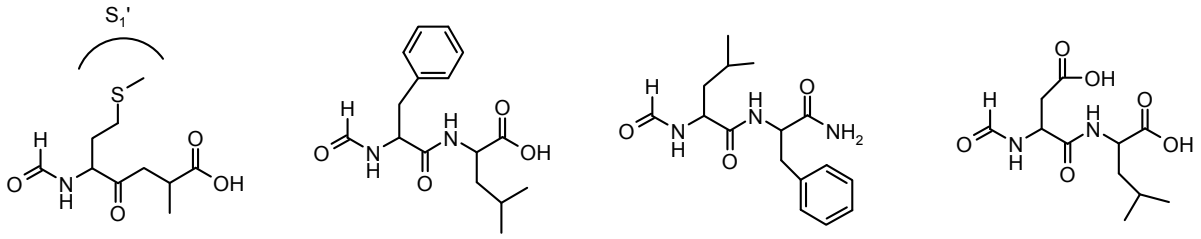


Figure 3. Reference substrate for PDF and selection substrates used in this study.

Determination of the specific activity of these 3 hits toward *N*-formyl-Phe-Leu-OH was carried out using CFE (see Materials and Methods) and revealed a 2.5-fold improvement compared to the $EcPDF_{Etag}$ wild type. Sequencing disclosed that all three carried asingle mutation C138V, each encoded by a different codon.

The same selection method was applied to identify mutants with activity towards *N*-formyl-Leu-Phe-NH_2. In this case, however, the negative control (non-mutated $EcPDF_{Etag}$ in *E. coli* MC1061) already demonstrated significant growth, precluding identification of improved mutants. A reduction of the L-arabinose concentration from 0.2 to 0.02 wt. % significantly reduced the wild-type growth and thus allowed discrimination between wild-type and improved mutants. Plating of about 350,000 CFUs on this medium resulted in the appearance of 44 colonies. The positive colonies were retested for their ability to grow on selective medium and 25 of them were confirmed. Subsequently, plasmid-encoded activity was proven for 4 of these, using a second selection round on the same selective medium containing *N*-formyl-Leu-Phe-NH_2 as only source of leucine. An increase in activity of about 2-fold toward the selection substrate was determined after preparation of the CFE from these clones. All 4 mutants appeared to contain the C138A mutation, encoded by the triplets GCC and GCG.

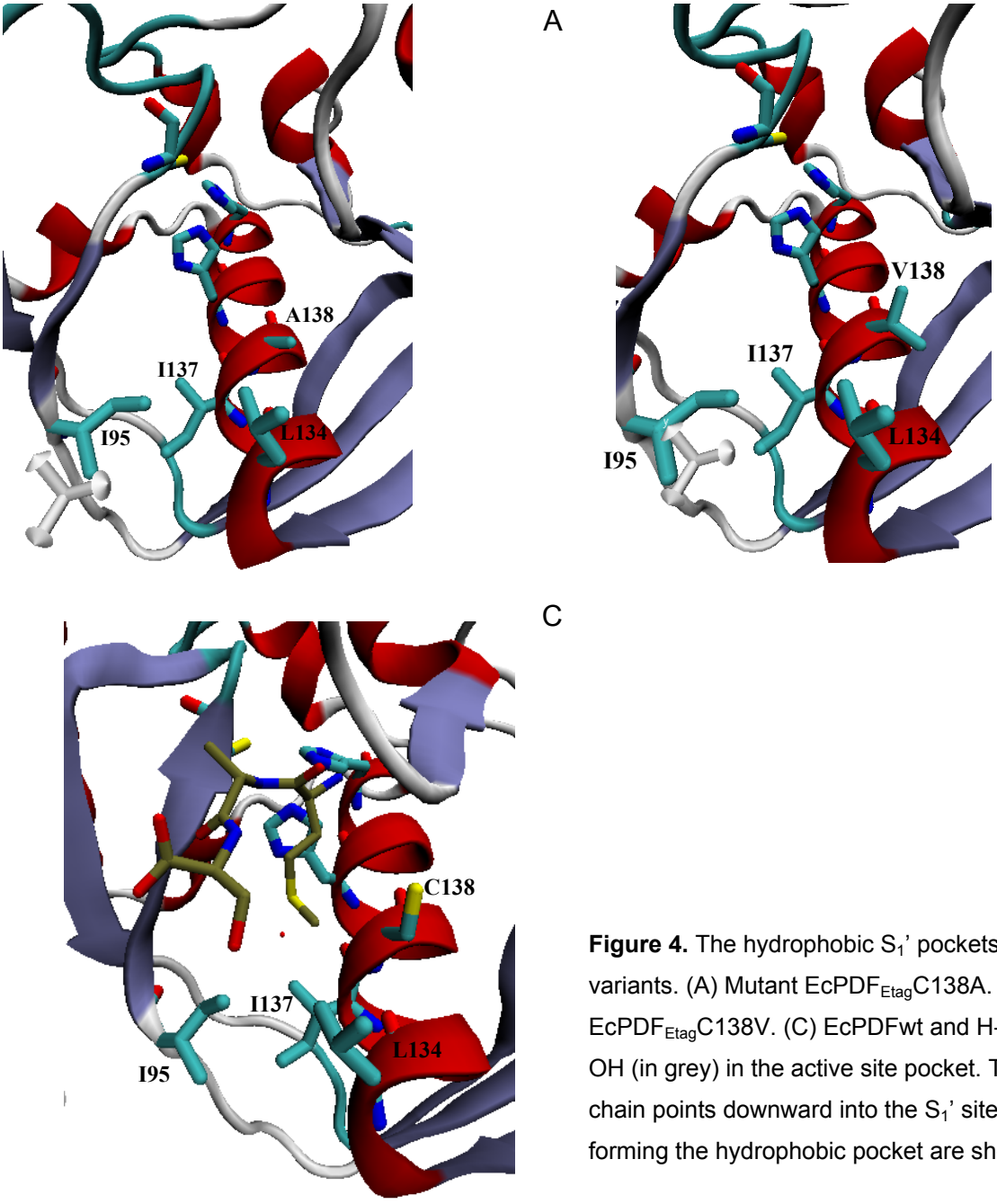


Figure 4. The hydrophobic S_1' pockets of EcPDF variants. (A) Mutant $EcPDF_{Etag}C138A$. (B) Mutant $EcPDF_{Etag}C138V$. (C) EcPDFwt and H-Met-Ala-Ser-OH (in grey) in the active site pocket. The Met side chain points downward into the S_1' site. Residues forming the hydrophobic pocket are shown in sticks.

As described above, the selection for *E. coli* clones that express a PDF variant that has acquired activity with *N*-formyl-Phe-Leu-OH or *N*-formyl-Leu-Phe-NH_2 was done on plates containing 0.2 and 0.02 % L-arabinose, respectively. The lower concentration of inducer was needed to suppress the growth of the non-mutated $EcPDF_{Etag}$ when using the latter substrate. This was surprising since the specific activity of $EcPDF_{Etag}$ is about 3 times higher for *N*-formyl-Phe-Leu-OH than for *N*-formyl-Leu-

Phe-NH_2. To further study why the $EcPDF_{Etag}$ allowed better growth with *N*-formyl-Leu-Phe-NH_2 than with *N*-formyl-Phe-Leu-OH (despite the higher catalytic efficiency of EcPDFfor the latter substrate) the activity of intact *E. coli* cells expressing $EcPDF_{Etag}$ wild type was determined with both formylated peptides and compared to that of the same cells towards the deformylated dipeptides. The results made clear that whole cells metabolize *N*-formyl-Leu-Phe-NH_2 faster than the other formylated substrate (Table 3). For example, 0.045 mg/l of *E. coli* MC1061 cells (wet weight) expressing $PDF_{E\text{-}tag}$ converted in 15 min 96% of a 1 mM *N*-formyl-Leu-Phe-NH_2 solution, whereas under the same conditions only 13% conversion was obtained with 1 mM *N*-formyl-Phe-Leu-OH. Both substrates were effectively split into separate amino acids after deformylation. Thus, the relative deformylation activity of whole cells for these two substrates is inverted as compared to that of free enzyme. Moreover the deformylated peptides are converted faster than the formylated ones. We can therefore conclude that the unexpected better growth on minimal medium with *N*-formyl-Leu-Phe-NH_2 as sole source of leucine can be traced back to a higher rate of substrate uptake by the *E. coli* cells.

Table 3. Degradation of *N*-formyl-Leu-Phe-NH_2 and *N*-formyl-Phe-Leu-OH by whole cells of *E. coli* MC1061 expressing PDF_{Etag}.

	Relative amount (area %)[b]	
Substrate	0.9 g[a]	0.09 g[a]
Exp 1 :		
N-formyl-Leu-Phe-NH_2	4	60
H-Phe-OH	73	26
H-Phe-NH_2	22	10
H-Leu-Phe-NH_2	1	4
Exp 2:		
N-formyl-Phe-Leu-OH	87	93
H-Phe-OH	12	7
H-Phe-Leu-OH	2	1

[a] Amount of wet cells used for activity determination (in 20 ml reaction volume).Cells were mixed with substrates (*N*-formyl-Leu-Phe-NH_2 or *N*-formyl-Phe-Leu-OH) and concentration of substrate was and reaction products were measured after 15 min.

[b] Percentage of substrate remaining and reaction products formed after incubation of whole cells with the substrate indicated (1 mM).

The third substrate used in the selection assay was *N*-formyl-Asp-Leu-OH. Despite the use of different L-arabinose concentrations for induction and a prolonged incubation time of the selection plates (up to 25 days), no colonies were observed. An explanation for this may be the very low catalytic efficiency of EcPDF wild-type for formylated dipeptides with an aspartic acid residue in position P_1'. Since the catalytic efficiency for this substrate is more than 100,000-fold lower than for a formylated peptide with an *N*-terminal methionine residue (see Table 1), we can speculate that the increase in activity needed for growth was just too high to be obtained in a single round of mutagenesis.

Partial characterization of Cys-138 EcPDFEtag mutants

Based on the observations in the primary screening that only mutations at position Cys-138 gave an increased activity towards the selection substrates, our interest shifted to a deeper understanding of the role Cys-138 plays in substrate recognition. This residue is positioned at the floor of the cavity and it is interacting with the methionine side chain of the substrate (Figure 4). We constructed all possible mutants at this position and tested the expression and activity of *E. coli* cells harboring these $EcPDF_{Etag}$ variants using gel electrophoresis and the activity assay described in Materials and Methods. The results of these experiments are presented in a qualitative mode in Table 3, and the gels are shown in Figure 5.

In total six substitutions at position Cys-138 (C138I, C138L, C138T, C138V, C138A and C138Y) led to expression of active enzyme. Using CFE, a first evaluation of their activity on several substrates showed that the variants posses a remarkable difference in substrate specificity. The nonpolar aliphatic side chains of alanine, valine and isoleucine at position 138 allow activity on several substrates, whereas the polar side chains of tyrosine and threonine give a lower activity as compared to the EcPDF wild-type towards all tested substrates.

The six active mutants were all overexpressed in *E. coli* TOP 10 and appeared stable. They were subsequently purified via ion exchange chromatography in order to determine, in a quantitative way, the substrate specificity towards a series of *N*-formylated peptides (Figure 5). Overexpression was high: about 10-20% of the total protein in cell-free extract of mutants C138I, C138L, C138T and C138V consisted of PDF, while in the case of C138Y and C138A the expression level was less than 10%.

After purification all mutants (Figure 6) showed a lower activity for *N*-formyl-Met-Ala-OH compared to EcPDF wild-type (Table 5). The activities found with pure enzyme confirmed the results of the growth selection. In fact, the C138V mutant, which was found with *N*-formyl-Phe-Leu-OH as selection substrate, showed the highest activity towards this substrate, with a 2.4-fold improvement as compared to the wild-type enzyme. The mutant was also highly active towards another substrate with a phenylalanine in position P1': *N*-formyl-Phe-Phe-NH_2, attesting the correlation between the mutation and the acceptance of the phenylalanine side chain in the modified S1' pocket. Moreover, the fact that mutant C138I was not found during the selection assay can be explained by the observation that its activity is similar to that of wild-type EcPDF.

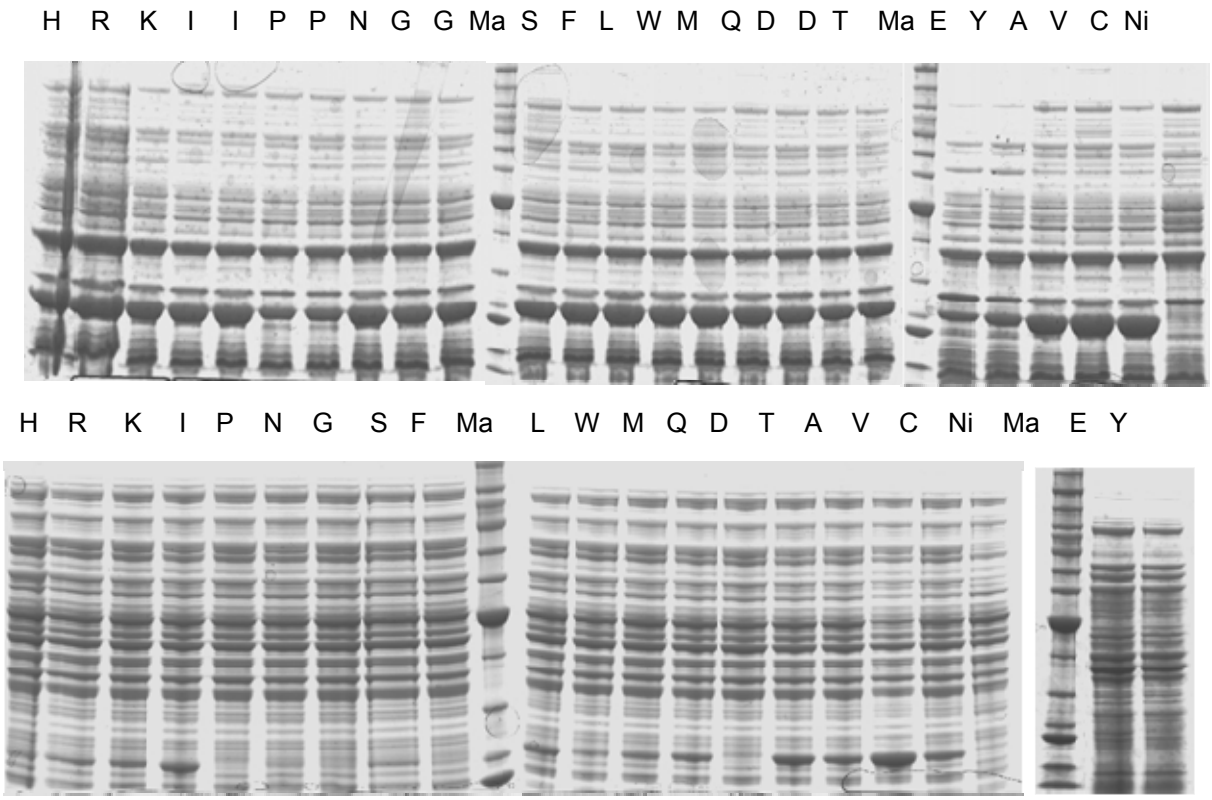


Figure 5. SDS-PAGE of the $EcPDF_{E\text{-}tag}$ mutants. Upper panel: whole cells. Lower panel: SDS soluble fraction. Letters indicate the C138 substitution introduced at position C138 in E. coli PDFE-tag. Ma: marker proteins, Ni, non-induced.

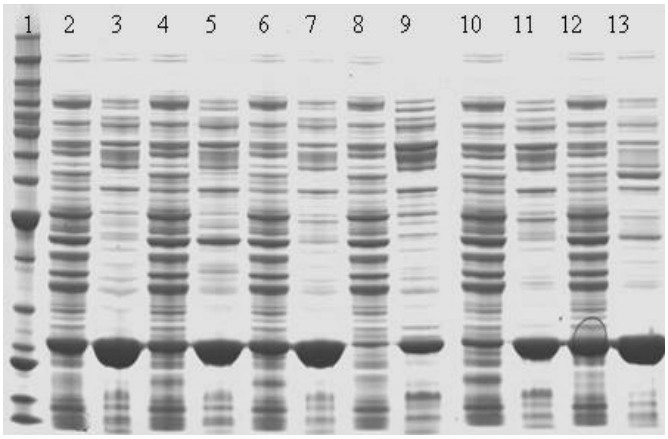
Table 4. Expression and substrate range of $EcPDF_{Etag}$ variants mutated at position 138.

Amino acid at position 138	Cys	His	Arg	Lys	Ile	Pro	Asn	Gly	Ser	Phe	Leu	Trp	Met	Gln	Asp	Thr	Glu	Tyr	Ala	Val
Expression[c]	++	+	+	+	++	−	−	−	+	−	++	+	+	++	−	++	−	−	++	++
Substrates/ activity																				
N-formyl-Met-Ala-OH	++	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	++
N-formyl-Leu-Phe-NH_2	++	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+++	++
N-formyl-Phe-Leu-OH	++	−	−	−	+++	−	−	−	−	−	++	−	−	−	−	+	−	+	++	+++
N-formyl-Val-Val-OH	++	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	++	++	+++
N-formyl-Lys-Phe-NH_2	++	−	+++	−	++	++	−	−	−	−	++	−	−	++	−	++	−	+	−	++
N-formyl-Phe-Phe-NH_2	++	−	−	−	+++	−	−	−	−	−	++	−	−	−	−	+	−	+	++	+++
N-formyl-Tyr-Leu-Phe-NH_2	++	−	−	−	+++	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−

[a] All assays were performed using CFE.

[b] Activities were scored as follows: −, < 5% conversion relative to $EcPDF_{Etag}$, +, 5-50% relative conversion; ++, 50-100% relative conversion; +++, > 100% relative conversion.

[c] Expression judged by SDS-PAGE of CFE (see Fig. 5)..



lane 1, Protein standard
lane 2, CFE C138I
lane 3, C138I purified
lane 4, CFE C138L
lane 5, C138L purified
lane 6, CFE C138T
lane 7, C138T purified
lane 8, CFE C138Y
lane 9, C138Y purified
lane 10, CFE C138A
lane 11, C138A purified
lane 12, CFE C138V
lane 13, C138V purified

Figure 6. SDS-PAGE of the $EcPDF_{Etag}$ mutants before and after purification via ion exchange chromatography; Lanes were loaded with 25 µg protein.

A similar evaluation was done considering *N*-formyl-Leu-Phe-NH_2 as substrate. In this case, the only mutant with a significantly increased activity towards the dipeptide is C138A. Again, the result of the activity assays and the growth selection matched. Furthermore the outcome of the retest demonstrated the strength of the growth selection assay. Thanks to a careful tuning of the conditions chosen for the selection using *N*-formyl-Leu-Phe-NH_2 we were able to discriminate between C138A and wild-type, having only 1.5-fold difference in activity (Table 5).

The largest improvement in substrate specificity was reached for the substrate *N*-formyl-Phe-Phe-NH_2. Clones C138V and C138I showed respectively a 7.7- and a 7.5-fold increased activity as compared to the wild type (Table 5). Despite the good results obtained, the activity towards the tested substrates still appeared much lower than the one measured with the natural substrate mimic, *N*-formyl-Met-Ala-OH.

We conclude that because of the prominent role played by residue Cys-138 in substrate recognition, most mutations at this position are extremely detrimental for the activity of $EcPDF_{Etag}$ with the natural substrate. Analysis of the production of the PDFs mutated at position 138 revealed formations of inclusion bodies in the cells, which were visible upon SDS-PAGE of pellets obtained by centrifugation of total cell lysates (data not shown). Inclusion body formation coincided with a low level of PDF in the soluble fraction and by an almost complete lack of deformylation activity, except in the case of

substitution of Cys-138 with isoleucine, alanine or valine. The aggregation of insoluble misfolded protein reveals the importance of position 138 for correct folding of $EcPDF_{Etag}$. According to previous studies conducted by Meinnel and coworkers [20], the substitution of Cys-129 in $EcPDF_{wt}$ (corresponding to Cys-138 in $EcPDF_{Etag}$) had very different effects on the catalytic efficiency. It was shown that C129W resulted in a reduction of hydrolysis rate by four orders of magnitude while with C129A the deformylation rate was identical to that of the wild-type enzyme. In contrast substitution of Ile-86, Ile-128 and of Leu-125 caused a decrease of catalytic efficiency in all cases although the remaining rate varied of up to one order of magnitude [20]

It is likely that for further enhancement of the activity of PDF towards different *N*-formylated dipeptides, more mutagenesis rounds that include other sites than Cys-138 would be needed. Future mutagenesis studies can identify other crucial residues for substrate recognition and simultaneous mutagenesis of these hotspots could lead to an altered substrate binding pockets with more dramatic effects on the substrate specificity of the enzyme.

Table 5. Peptide deformylase activity of $EcPDF_{Etag}$ and different C138 mutants purified via ion exchange chromatography.

Compound	Sp. Activity (U/mg) of:						
	$EcPDF_{Etag}$	C138I	C138L	C138T	C138Y	C138A	C138V
N-formyl-Met-Ala-OH	970	140	30	140	0	420	770
N-formyl-Val-Val-OH	0.4	0.2	0	0	0.2	0.4	0.8
N-formyl-Phe-Leu-OH	12	13	3	2	0	5	30
N-formyl-Lys-Phe-NH_2	0.03	0	0	0.04	0	0.01	0.01
N-formyl-Leu-Phe-NH_2	4	0	0	2	0	6	2
N-formyl-Phe-Phe-NH_2	4	30	5	1	2	2	31
N-formyl-Tyr-Leu-Phe-NH_2	4	10	3	1	0	1	12

Activity was measured using 0.6 mM substrate at pH 7.2.

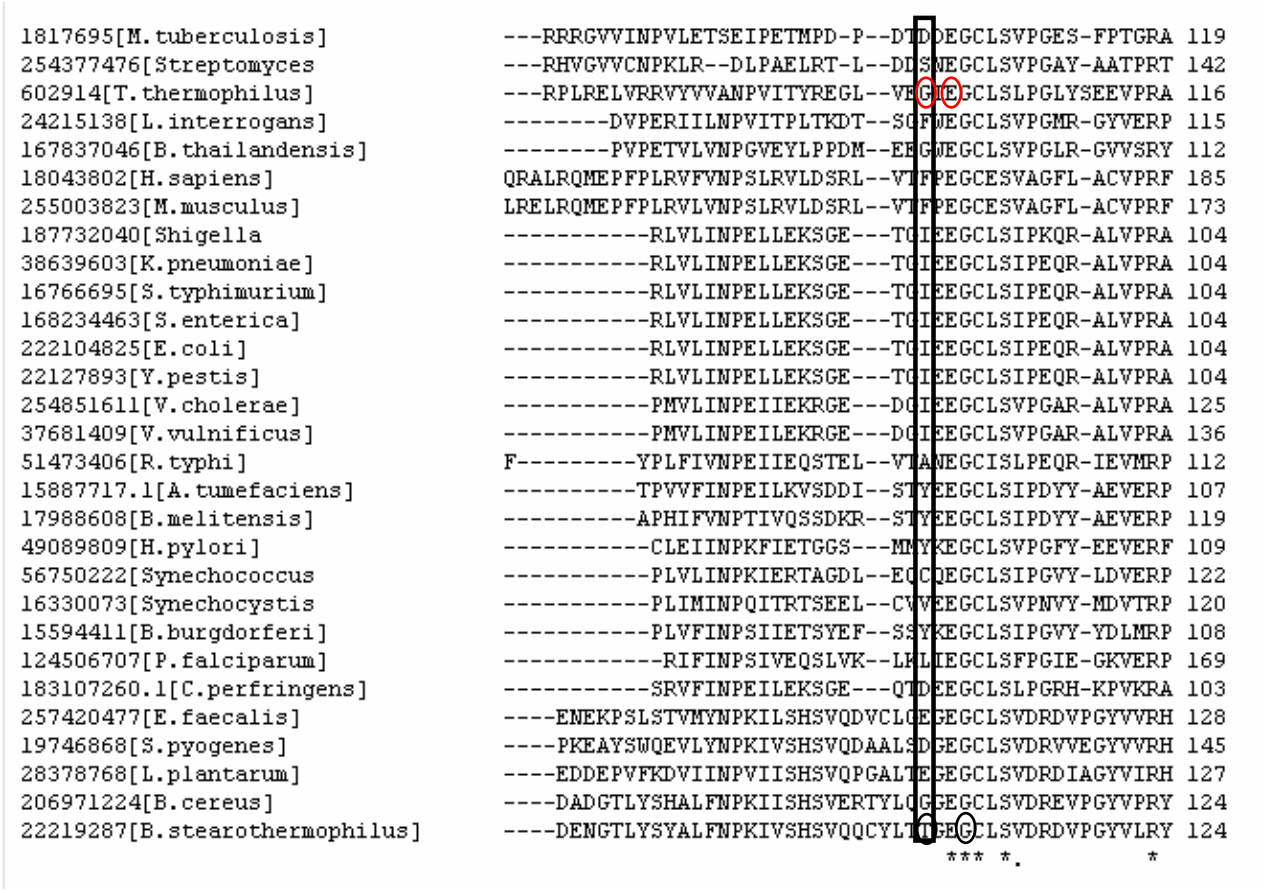

1817695[M.tuberculosis] ---RRRGVVINPVLETSEIPETMPD-P--DTDQEGCLSVPGES-FPTGRA 119
254377476[Streptomyces ---RHVGVVCNPKLR--DLPAELRT-L--DLSNEGCLSVPGAY-AATPRT 142
602914[T.thermophilus] ---RPLRELVRRVYVVANPVITYREGL--VEGIEGCLSLPGLYSEEVPRA 116
24215138[L.interrogans] --------DVPERIILNPVITPLTKDT--SGFWEGCLSVPGMR-GYVERP 115
167837046[B.thailandensis] --------PVPETVLVNPGVEYLPPDM--EEGWEGCLSVPGLR-GVVSRY 112
18043802[H.sapiens] QRALRQMEPFPLRVFVNPSLRVLDSRL--VTFPEGCESVAGFL-ACVPRF 185
255003823[M.musculus] LRELRQMEPFPLRVLVNPSLRVLDSRL--VTFPEGCESVAGFL-ACVPRF 173
187732040[Shigella ----------RLVLINPELLEKSGE---TGIEEGCLSIPKQR-ALVPRA 104
38639603[K.pneumoniae] ----------RLVLINPELLEKSGE---TGIEEGCLSIPEQR-ALVPRA 104
16766695[S.typhimurium] ----------RLVLINPELLEKSGE---TGIEEGCLSIPEQR-ALVPRA 104
168234463[S.enterica] ----------RLVLINPELLEKSGE---TGIEEGCLSIPEQR-ALVPRA 104
222104825[E.coli] ----------RLVLINPELLEKSGE---TGIEEGCLSIPEQR-ALVPRA 104
22127893[Y.pestis] ----------RLVLINPELLEKSGE---TGIEEGCLSIPEQR-ALVPRA 104
25485161l[V.cholerae] ----------PMVLINPEIIEKRGE---DGIEEGCLSVPGAR-ALVPRA 125
37681409[V.vulnificus] ----------PMVLINPEILEKRGE---DGIEEGCLSVPGAR-ALVPRA 136
51473406[R.typhi] F---------YPLFIVNPEIIEQSTEL--VTAWEGCISLPEQR-IEVMRP 112
15887717.1[A.tumefaciens] ----------TPVVFINPEILKVSDDI--STYEEGCLSIPDYY-AEVERP 107
17988608[B.melitensis] ----------APHIFVNPTIVQSSDKR--STYEEGCLSIPDYY-AEVERP 119
49089809[H.pylori] ----------CLEIINPKFIETGGS---MMYKEGCLSVPGFY-EEVERF 109
56750222[Synechococcus ----------PLVLINPKIERTAGDL--EQCQEGCLSIPGVY-LDVERP 122
16330073[Synechocystis ----------PLIMINPQITRTSEEL--CWEEGCLSVPNVY-MDVTRP 120
15594411[B.burgdorferi] ----------PLVFINPSIIETSYEF--SSYKEGCLSIPGVY-YDLMRP 108
124506707[P.falciparum] ----------RIFINPSIVEQSLVK--LKLIEGCLSFPGIE-GKVERP 169
183107260.1[C.perfringens] ----------SRVFINPEILEKSGE---QTDEEGCLSLPGRH-KPVKRA 103
257420477[E.faecalis] ----ENEKPSLSTVMYNPKILSHSVQDVCLQEGEGCLSVDRDVPGYVVRH 128
19746868[S.pyogenes] ----PKEAYSWQEVLYNPKIVSHSVQDAALSDGEGCLSVDRVVEGYVVRH 145
28378768[L.plantarum] ----EDDEPVFKDVIINPVIISHSVQPGALTEGEGCLSVDRDIAGYVIRH 127
206971224[B.cereus] ----DADGTLYSHALFNPKIISHSVERTYLQGGEGCLSVDREVPGYVPRY 124
22219287[B.stearothermophilus] ----DENGTLYSYALFNPKIVSHSVQQCYLTGEGCLSVDRDVPGYVLRY 124
*** *. *
1817695[M.tuberculosis] KWARVTGLDADGSPVSIEGTGLFARMLQHETGHLDGFLYLDRLIGRYARN 169
254377476[Streptomyces EYAEVTGQDEHGNAIKVRGTGYFARCLQHETDHLYGSLYLDRLS---KRD 189
602914[T.thermophilus] ERIRVEYQDEEGRGRVLELEGYMARWFQHEIDHLDGILFFERLP---KPK 163
24215138[L.interrogans] NQIRMQWMDEKGNQFDETIDGYKAIVWQHECDHLQGILYVDRLKDTKLFG 165
167837046[B.thailandensis] RRVRYSGFDQYGAKLERIAEGFHARVWQHEYDHLIGKLYPMRITDFSKFG 162
18043802[H.sapiens] QAVQISGLDPNGEQVVWQASGWAARIIQHEMDHLQGCLFIDKMDSR---T 232
255003823[M.musculus] QAVQISGLDPKGEPVVWSASGWTARIIQHEMDHLQGCLFIDKMDSG---T 220
187732040[Shigella EKVKIRALDRDGKPFELEADGLLAICIQHEMDHLVGKLFMDYLSPLKQQR 154
38639603[K.pneumoniae] EKVKIRALNRYGKPFELEADGLLAICIQHEMDHLVGKLFMDYLSPLKQQR 154
16766695[S.typhimurium] EKVKIRALDRDGNPFELEADGLLAICIQHEMDHLVGKLFIDYLSPLKQQR 154
168234463[S.enterica] EKVKIRALDRDGNPFELEANGLLAICIQHEMDHLVGKLFIDYLSPLKQQR 154
222104825[E.coli] EKVKIRALDRDGKPFELEADGLLAICIGLRLGN--GK----YCT------ 142
22127893[Y.pestis] EKVKIRALDRDGKPFELETDGLLAICIQHEMDHLIGKLFVDYLSPLKRQR 154
25485161l[V.cholerae] AEVTVKALDRNGQEYQFDADDLLAICVQHELDHLAGKLFVDYLSPLKRNR 175
37681409[V.vulnificus] AEVTVKALDRDGHEFTLEADDLLAICIQHELDHLQGKLFVDYLSPLKRKR 186
51473406[R.typhi] ESVKIRYLDYHGKSQELKANDWLARVIQHEYDHLEGKLMVDYLSNLKRDV 162
15887717.1[A.tumefaciens] ASLTVQYVGRDGKQQTVEADGLLATCLQHEIDHLNGVLFIDHISRLKRDM 157
17988608[B.melitensis] ATVKVNYFDADGKPQSMEADGLMATCLQHEIDHLNGVLFIDHISKLKRDM 169
49089809[H.pylori] EKVKIEYQNRFAEVKVLEASELLAVAIQHEIDHLNGVLFVDKLSILKRKK 159
56750222[Synechococcus EIVEVSYKDENGRPQRLVADGLLARCIQHEMDHLNGVLFVDRVENRLELN 172
16330073[Synechocystis RAIEVTYKDEHGRPQKRLFARLTARVIQHEMDHLNGVMFVDRVDNPLALA 170
15594411[B.burgdorferi] KAVVINFHDENGKSFTIENSDFLARIIQHEMDHLNGVLFIDYYEEKLKN- 157
124506707[P.falciparum] SIVSISYYDINGYKHLKILKQIHSRIFQHEFDHLNGTLFIDKMTQVDKKK 219
183107260.1[C.perfringens] NKIKIKALDVNGNEFVLDAEGFLARAIQHEYDHLEGVLFIDHEL------ 147
257420477[E.faecalis] NKITVSYFDMAGEKHKVRLKNYEAIVWQHEIDHINGIMFYDHINKENPFA 178
19746868[S.pyogenes] ARVTVDYYDKEGQQHRIKLKQYNAIVWQHEIDHINGILFYDRINAKNPFE 195
28378768[L.plantarum] DRITLRYYNMAGEEKKIRLKNYPAIVCQHEIDHLHGILFYDHINGDNPFA 177
206971224[B.cereus] TRITVKATSINGEEIKLRLKQLPAIVFQHEIDHLNGVMFYDHINKENPFA 174
22219287[B.stearothermophilus] ARITVTGTTLDGEEVTLRLKQLPAIVFQHEIDHLNGIMFYDRINPADPFQ 174
. : . .: *

Figure 7. Alignment of the $EcPDF_{wt}$ with the sequences of twenty-nine other peptide deformylases from different organisms. The alignment was generated with ClustalW2 at the server of the European Bioinformatics Institute. The enzyme codes refer to their accession numbers. The residues highlighted by the boxes indicate the sites mutated in this study. The residues marked with a circle line the S1' pocket of BsPDF and TtPDF.

Mutation of the S_1' pocket

Comparisons of structures and sequences of S1' pockets of PDFs from different organisms using alignment programs and visualization software revealed that the variable residues found in the S1' cleft modify its overall shape only slightly, while retaining the hydrophobic character that is responsible for the affinity to and anchoring of the substrate Results of alignments show that among the non-conserved residues, at least two are changed simultaneously as compared to the *E. coli* PDF, suggesting a change in the pocket shape while maintaining the hydrophobic character, which will allow the methionine side chain still to fit when it adopts a different conformation [9].. The alignment showed that the S1' pocket of *B. stearothermophilus* (BsPDF) is formed by Glu-108 and His-153 delineating the ceiling, Val-150 and Ile-59 at the floor and Ile-149, Thr-106 and Leu-146 framing the back. In the case of *T. thermophilus* (TtPDF) the residues are Glu-100 and His-145 for the ceiling, Val-43 and Val-142 for the floor and Arg-141, Tyr-138 and Gly-98 composing the back (Figure 7). Our attention was therefore pointed at the presence of valine instead of cysteine at a position that corresponds to Cys-138 of *E. coli* PDF_{Etag}.

Because of the important role played by Cys-138 in defining the substrate specificity of EcPDF, the activity of BsPDF and TtPDF, which carry a valine at the corresponding position in the S_1' cleft, was tested against *N*-formyl-Phe-Leu-OH. The results showed no elevated activity of these PDFs (Table 6). From a structural analysis of the hydrophobic pockets of BsPDF and TtPDF, we noticed that several residues are different compared to $EcPDF_{Etag}$. For example, in TtPDF Arg-141, orienting its side chain towards the back of the pocket is present instead of the Ile-137 of $EcPDF_{Etag}$, and Tyr-138 is closing the bottom of the pocket. In the case of BsPDF, Thr-106 is replacing Ile-95 of $EcPDF_{Etag}$ while the other residues of the pocket are conserved. Despite this, the orientation of their side chains is different compared to EcPDF (Figure 7), probably leading to a different orientation of the *N*-formylated peptide in the active site compared to EcPDF, which still allows high catalytic efficiency when methionine is present in

position P_1 but does not cause an increased activity towards other hydrophobic amino acid residues.

Table 6. Peptide deformylase activity in *E. coli* TOP10 CFE overexpressing $EcPDF_{Etag}$, *B. stearothermophilus* PDF (BsPDF) and *T. thermophilus* PDF (TtPDF).

	Sp. Activity (U/mg)		
Compound	$EcPDF_{Etag}$	BsPDF	TtPDF
N-formyl-Met-Ala-OH	12	17	14
N-formyl-Phe-Leu-OH	0.6	0.1	0.2

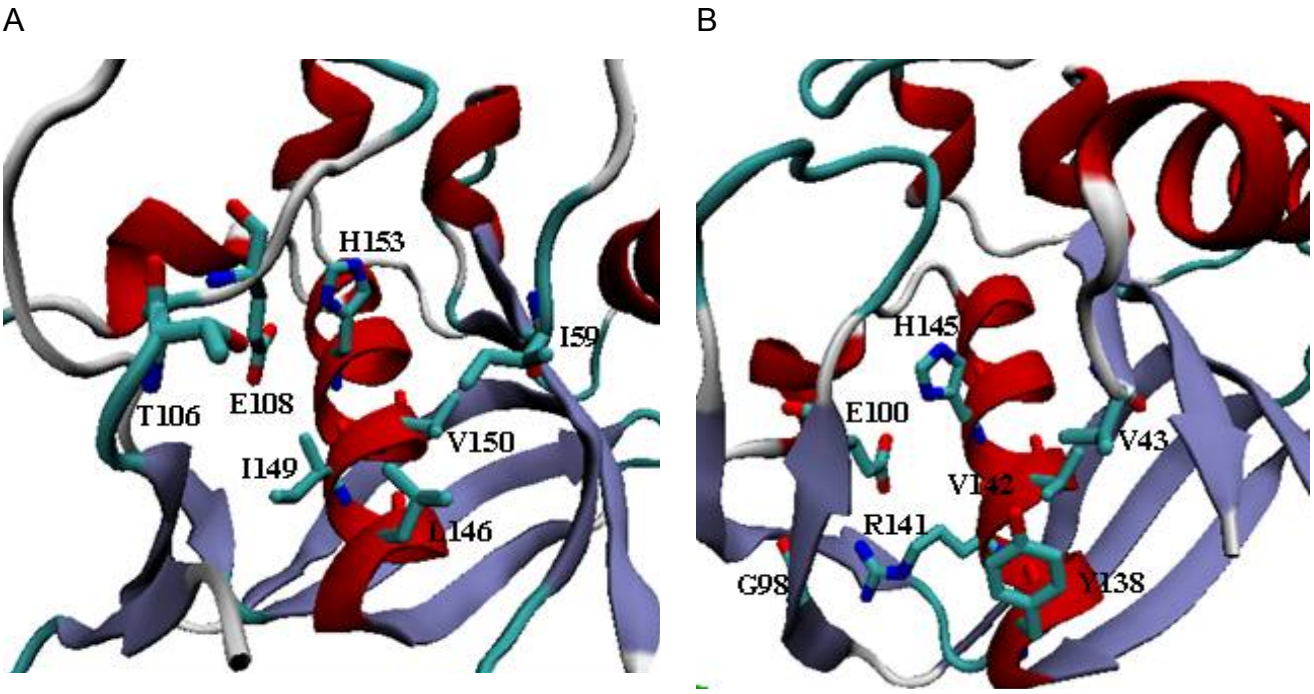


Figure 7. Representations of the hydrophobic pockets of BsPDF (A) and $TtPDF_{Etag}$ (B). Residues forming the hydrophobic pocket are shown in sticks.

Previous studies on *E. coli* PDF showed this enzyme to have a relatively high affinity for hydrophobic amino acid residues other than methionine (as indicated by their low K_M value) but also a very low hydrolysis rate (k_{cat}). This is possibly due to a wrong orientation of the *N*-formyl group in the catalytic site [7].

From the mutagenesis work on the $EcPDF_{Etag}$ it appears that very few mutations are allowed in order to maintain activity. This is in contrast with the fact that PDF from

other organisms present a different set of amino acids around the S_1' pocket and yet are highly active towards the natural substrate. We can therefore conclude that it is not just the change of one amino acid that can strongly influence the substrate specificity, but a combination of substitutions that enables a change of the overall architecture of the cleft.

Based on the marked substrate specificity of peptide deformylases and on the obtained results, it is tempting to assume that Cys-138 in combination with the other residues delineating the pocket is evolutionary the best residue for activity towards the natural substrate. In fact, a synergy among the variable residues delineating the S1' cleft appears to be important for retaining an overall hydrophobic character and substrate specificity.

Conclusions

The mutagenesis data presented here show that very few mutations are tolerated in the hydrophobic S_1' pocket that binds the methionine side chain of *E. coli* PDF. From a library that is randomized at four non-conserved positions that surround this pocket, few active variants were discovered and they all carried mutations only at residue Cys-138. These variants showed elevated activity with target *N*-formyl-peptides that were used for selection, demonstrating that it is indeed possible to increase the activity of a PDF using the approach described here.

The analysis of a library of mutants at position 138 confirmed that only a few mutations are allowed, with most substitutions giving either unfolded (aggregated) enzyme or enzyme with very low activity. Thus, it appears difficult to change the selectivity while maintaining activity of *E. coli* PDF by mutating residues around the S1' pocket only. The fact that the mutated residues are not conserved in the sequence of homologous PDFs suggest that compensating mutations are required in residues that do not directly surround the active site, but instead occupy positions in the second shell around the active site. A strategy to maintain activity while changing substrate specificity could thus be to modify a larger number of second shell residues simultaneously with the amino acids that surround the S_1' pocket.

References

1. **Antikainen, N. M., P. J. Hergenrother, M. M. Harris, W. Corbett, and S. F. Martin**. 2003. Altering substrate specificity of phosphatidylcholine-preferring phospholipase C of *Bacillus cereus* by random mutagenesis of the headgroup binding site. Biochemistry **42**:1603-1610.
2. **Baldwin, E. T., M. S. Harris, A. W. Yem, C. L. Wolfe, A. F. Vosters, K. A. Curry, R. W. Murray, J. H. Bock, V. P. Marshall, J. I. Cialdella, M. H. Merchant, G. Choi, and M. R. Deibel, Jr.** 2002. Crystal structure of type II peptide deformylase from *Staphylococcus aureus*. J. Biol. Chem. **277**:31163-31171.
3. **Bershtein, S. and D. S. Tawfik**. 2008. Advances in laboratory evolution of enzymes. Curr. Opin. Chem. Biol. **12**:151-158.
4. **Bovenberg, R. and Kerkman, R**. 2003. Process for preparing variant polynucleotides. US Patent 7402383.
5. **Chica, R. A., N. Doucet, and J. N. Pelletier**. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr. Opin. Biotechnol. **16**:378-384.
6. **Dardel, F., S. Ragusa, C. Lazennec, S. Blanquet, and T. Meinnel**. 1998. Solution structure of nickel-peptide deformylase. J. Mol. Biol. **280**:501-513.
7. **Dirk, L. M. A., J. J. Schmidt, Y. Cai, J. C. Barnes, K. M. Hanger, N. R. Nayak, M. A. Williams, R. B. Grossman, R. L. Houtz, and D. W. Rodgers**. 2008. Insights into the substrate specificity of plant peptide deformylase, an essential enzyme with potential for the development of novel biotechnology applications in agriculture. Biochem. J. **413**:417-427.
8. **Fieulaine, S., C. Juillan-Binard, A. Serero, F. Dardel, C. Giglione, T. Meinnel, and J. L. Ferrer**. 2005. The crystal structure of mitochondrial (type 1A) peptide deformylase provides clear guidelines for the design of inhibitors specific for the bacterial forms. J. Biol. Chem. **280**:42315-42324.
9. **Guilloteau, J. P., M. Mathieu, C. Giglione, V. Blanc, A. Dupuy, M. Chevrier, P. Gil, A. Famechon, T. Meinnel, and V. Mikol**. 2002. The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: a platform for the structure-based design of antibacterial agents. J. Mol. Biol. **320**:951-962.
10. **Hu, Y. J., Y. Wei, Y. Zhou, P. T. R. Rajagopalan, and D. Pei**. 1999. Determination of substrate specificity for peptide deformylase through the screening of a combinatorial peptide library. Biochemistry **38**:643-650.
11. **Jank, T., D. J. Reinert, T. Giesemann, G. E. Schulz, and K. Aktories**. 2005. Change of the donor substrate specificity of *Clostridium difficile* toxin B by site-directed mutagenesis. J. Biol. Chem. **280**:37833-37838.
12. **Kataoka, K. and K. Tanizawa**. 2003. Alteration of substrate specificity of leucine dehydrogenase by site-directed mutagenesis. J. Mol. Catal. B: Enzymatic **23**:299-309.
13. **Khouri, H. E., T. Vernet, R. Menard, F. Parlati, P. Laflamme, D. C. Tessier, B. Gour-Salin, D. Y. Thomas, and A. C. Storer**. 1991. Engineering of papain: selective alteration of substrate specificity by site-directed mutagenesis. Biochemistry **30**:8929-8936.
14. **Kreusch, A., G. Spraggon, C. C. Lee, H. Klock, D. McMullan, K. Ng, T. Shin, J. Vincent, I. Warner, C. Ericson, and S. A. Lesley**. 2003. Structure analysis of peptide deformylases from *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Thermotoga*

maritima and *Pseudomonas aeruginosa*: snapshots of the oxygen sensitivity of peptide deformylase. J. Mol. Biol. **330**:309-321.

15. **Kumar, A., K. T. Nguyen, S. Srivathsan, B. Ornstein, S. Turley, I. Hirsh, D. Pei, and W. G. J. Hol**. 2002. Crystals of peptide deformylase from *Plasmodium falciparum* reveal critical characteristics of the active site for drug design. Structure **10**:357-367.
16. **Meinnel, T., C. Lazennec, S. Villoing, and S. Blanquet**. 1997. Structure-function relationships within the peptide deformylase family. Evidence for a conserved architecture of the active site involving three conserved motifs and a metal ion. J. Mol. Biol. **267**:749-761.
17. **Nam, K. H., K. H. Kim, E. E. K. Kim, and K. Y. Hwang**. 2009. Crystal structure of an *Ef*PDF complex with Met-Ala-Ser based on crystallographic packing. Biochem. Biophys. Res. Commun. **381**:630-633.
18. **Penning, T. M. and J. M. Jez**. 2001. Enzyme Redesign. Chem. Rev. **101**:3027-3046.
19. **Quaedflieg P., Sonke, T., and Wagner, V**. 2000. Process for the enzymatic preparation of amino acid derivatives with enhanced optical purity. Patent WO/2000/037657.
20. **Ragusa, S., P. Mouchet, C. Lazennec, V. Dive, and T. Meinnel**. 1999. Substrate recognition and selectivity of peptide deformylase. Similarities and differences with metzincins and thermolysin. J. Mol. Biol. **289**:1445-1457.
21. **Robien, M. A., K. T. Nguyen, A. Kumar, I. Hirsh, I. Turley, D. Pei, and W. G. J. Hol**. 2003. An improved crystal form of *Plasmodium falciparum* peptide deformylase. Protein Sci. **13**:1155-1163.
22. **Smith, K. J., C. M. Petit, K. Aubart, M. Smyth, E. McManus, J. Jones, A. Fosberry, C. Lewis, M. Lonetto, and S. B. Christensen**. 2003. Structural variation and inhibitor binding in polypeptide deformylase from four different bacterial species. Protein Sci. **12**:349-360.
23. **Sonke, T., B. Kaptein, A. F. V. Wagner, P. J. L. M. Quaedflieg, S. Schultz, S. Ernste, A. Schepers, J. H. M. Mommers, and Q. B. Broxterman**. 2004. Peptide deformylase as biocatalyst for the synthesis of enantiomerically pure amino acid derivatives. J. Mol. Catal. B: Enzymatic **29**:265-277.
24. **Walker, K. W. and R. A. Bradshaw**. 1999. Yeast methionine aminopeptidase I. Alteration of substrate specificity by site directed mutagenesis. J. Biol. Chem. **274**:13403-13409.
25. **Whittle, E. and J. Shanklin**. 2001. Engineering $\Delta^{9\text{-}16:0\text{-}Acyl}$ carrier protein (ACP) desaturase specificity based on combinatorial saturation mutagenesis and logical redesign of the castor $\Delta^{9\text{-}18:0\text{-}ACP}$ desaturase. J. Biol. Chem. **276**:21500-21505

Chapter 6

Summary and conclusions

Concluding remarks

Peptides built of α-, β- or γ-amino acids can have very different biological functions. For example, they can act as hormones, neuropeptides, antibiotics, toxins or they may be involved in immune protection. Due to these diverse effects on physiological processes peptides are recognized as having high potential in medical applications. Peptides are also increasingly important as health- and taste-enhancing food ingredients and they may improve nutritional value. The growing interest in the physiological role of peptides and their emerging applications in foods and medicine contribute to an increased demand for cost-efficient and versatile methodologies for peptide synthesis, as well as better procedures for their isolation and analysis.

Three main approaches are available for the preparation of small peptides: chemical synthesis, production by fermentatation, and chemo-enzymatic peptide synthesis. Despite the widespread use of fermentative production for large peptides and proteins and the availability of good methods for chemical synthesis of short peptides, the chemo-enzymatic approach is considered of growing importance, especially for the production of peptides shorter than 10 amino acids. The key step in this process, i.e. the coupling of amino acids to form peptide bonds, is based on enzymatic catalysis. It has the advantage of mild reaction conditions that prevent undesired racemisation. Moreover the selectivity of enzymes allows a limited need for protection/deprotection of amino acid side chains, which reduces costs [17;26].

The most useful enzymes for peptide bond synthesis are proteases that can be used in thermodynamically or kinetically controlled coupling reactions. In both cases a proper strategy for the introduction and removal of activating or protecting groups needs to be employed during the synthesis of complex molecules [2]. Despite the existence of hundreds of different protecting groups, the search for cheap and easily cleavable ones is still ongoing. An attractive solution would be the application of biocatalysis to carry out highly chemo- and regioselective deprotection. A cheap, enzymatically cleavable group like formyl (For) is industrially attractive since it can easily be introduced in a cost-effective manner at the *N*-terminus of peptides and it can be removed enzymatically. Formylation can be done using formic acid and acetic acid anhydride [18]. The work described in this thesis is focused on the application of peptide deformylase as biocatalyst for the removal of the formyl group.

In eubacteria, ribosome-mediated protein synthesis starts with the *N*-formylation of the methionyl moiety, that will become the first amino acid building block of a newly synthesised peptide [22]. After peptide synthesis, the formyl group is cleaved off from the growing peptide by the action of peptide deformylase. Next, methionyl aminopeptidase may remove the *N*-terminal methionine leading to a mature protein [5]. PDF activity is essential for the survival of eubacteria, but it is also important in eukaryotes because protein synthesis in eukaryotic organelles such as mitochondria also starts with *N*-formyl methionine [8;9;11;24]. Its importance for bacterial and mitochondrial protein synthesis and growth has stimulated extensive research on the possibility to use PDFs as a target for antibacterial and antiparasitic agents [9;15]. However, at the start of this work, data on biocatalytic applications of PDFs were scarce. Only one publication described the application of the enzyme in stereoselective synthesis of amines and amino acid derivatives and the use of PDF for the selective deformylation of *N*-formyl dipeptides [28].

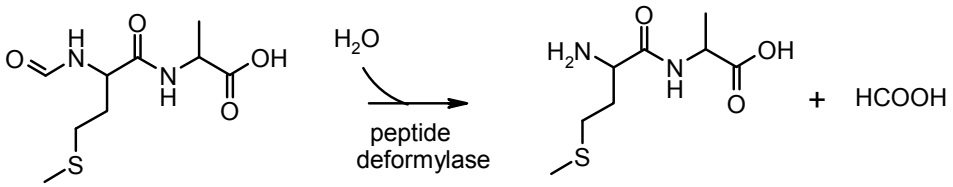


Figure 1. Peptide deformylase reaction with For-Met-Ala-OH.

The formyl moiety can be used as a protecting group that prevents a nucleophilic reaction of the N-terminal amino group of an amino acid or peptide that acts as the acyl donor in peptide-coupling reactions. If the formyl group is applied for this, it should be removed after the coupling step and prior to the next elongation step by a mild method that keeps the synthesized peptide bond intact and does not cause racemisation is needed for removal of the formyl group after completion of the coupling and prior to the next step. Acid-mediated chemical hydrolysis is the conventional method for the removal of formyl group from the *N*-terminus of peptides, but these harsh conditions lead to significant concomitant peptide bond hydrolysis.

The use of a bacterial peptide deformylase offers a milder way for this deprotection step. However, adequate purification steps are required for obtaining PDF preparations that are suitable for peptide deformylation, since extracts of a production organism like *E. coli* contain amidases and peptidases that cause extensive peptide bond hydrolysis. The work detailed in Chapter 2 explores the development of two simple

and industrially applicable purification methods for the PDF from *E. coli* (EcPDF). Because of the small size of EcPDF (molecular mass 19 kDa)[3] compared to the average size of *E. coli* amidases and peptidase [1], we first tested ultrafiltration as a simple and cost-effective way to separate EcPDF from amidases and peptidases present in the cell-free extract (CFE). In order to enhance the difference in size, we included testing of shortened version of EcPDF obtained by removal of its C-terminal α-helix at the gene level. The purification of both versions of EcPDF was attempted using membranes with different cut-off values (50 and 100 KDa) and filtration conditions. The passage of PDF through the membrane was heavily influenced by the size of the pores but also by protein concentration, pH, ionic strength and mechanical stress. The best purification results, leading to a recovery of 21% of the EcPDF activity in the filtrate, were obtained with a 100 kDa filter. The level of contaminating amidases in the purified sample decreased by 27-fold compared to the CFE, which was judged as being insufficient. A possible cause for this moderate performance is membrane fouling and clogging leading to poor flux and aspecific size discrimination. These phenomena can possibly be reduced by using tangential flow filtration since in this technique the shear flow reduces gel-like depositions on the membrane surface. Nevertheless when applying tangential flow for PDF purification, similar unsatisfactory results were obtained. Furthermore, the poor separation with filters with smaller pore sizes suggests formation of larger supramolecular complexes or aggregates including the EcPDF.

In parallel, a different purification approach was explored using ion-exchange chromatography in combination with the introduction of a negatively charged tag that was fused into a flexible and exposed loop of *E. coli* PDF. The tag was expected to enhance the binding of the engineered PDF to an ion-exchange column, increasing separation from background *E. coli* proteins, and providing a separation method that does not require expensive affinity material such as those used for hexahistidine- or glutathione transferase affinity tags. The overexpression of the octaglutamic acid-tagged PDF variant ($EcPDF_{Etag}$) was similar to that of the wild-type enzyme, while its specific activity was only slightly affected by the presence of the acidic tag. As a result, $EcPDF_{Etag}$ was purified from an overexpressing *E coli* strain by ion-exchange chromatography with 64% recovery of activity and a 1,300-fold decrease of the level of contaminating amidases.

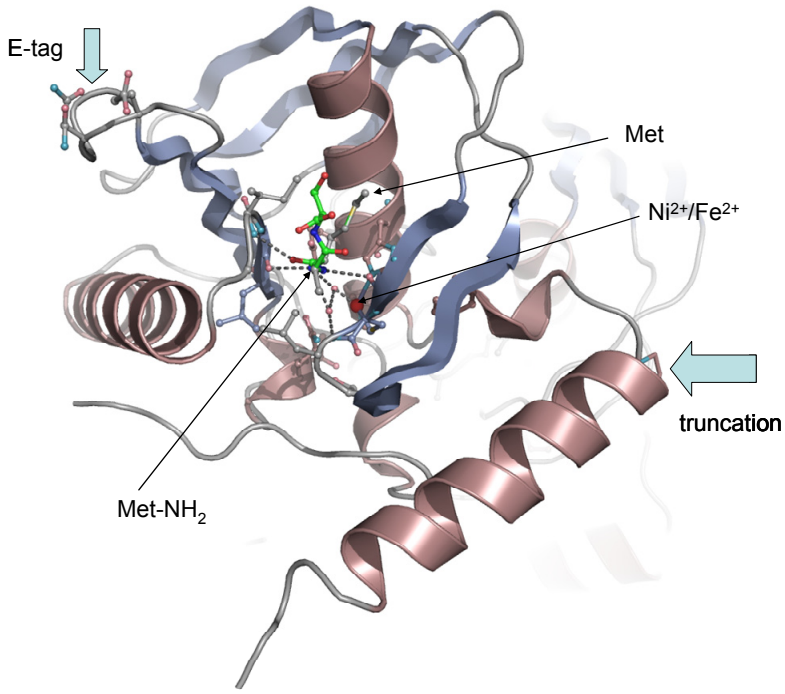


Figure 2. Structure of E-tag *E.coli* peptide deformylase. E-tag: place of insertion of the octaglutamate E-tag. Truncation: siteof proline, which with the C-terminal helix is removed in the short PDF variant. The methionine side chain and methionine amino group of the ligand H-Met-Ala-Ser-OH are also shown.

The purified octaglutamic acid-tagged enzyme appeared to hydrolyse *N*-formyl dipeptides without significant concomitant peptide bond hydrolysis. It was used for the chemo-enzymatic synthesis of the tripeptide *N*-formyl-Tyr-Leu-Phe-NH_2. Following the coupling of *N*-formyl-Leu with Phe-NH_2, the purified dipeptide was deprotected with $EcPDF_{Etag}$ resulting in the deformylated products in 80% isolated yield. The deprotected dipeptide was subsequently coupled to *N*-formylated tyrosine yielding the expected tripeptide. Thus, the enzyme equipped with the purification tag can be rapidly purified and used in biocatalytic conversions of industrial interest.

Important properties for judging the applicability of peptide deformylase in industrial biocatalysis are the stability and substrate specificity of the enzyme. Especially the requirement of a methionine or other hydrophobic group at the formylated N-terminal position of the peptide substrate is a serious limitation. A possible way to obtain a peptide deformylase with a relaxed substrate specificity is to explore natural biodiversity by genome mining. By overexpressing in *E. coli* peptide deformylase genes from different organisms, and subsequently characterize their properties,, we aimed at discovering more stable enzymes with more relaxed substrate specificities.

Various databases were screened for the presence of peptide deformylase-encoding sequences, using the sequence of EcPDF as a query. In total, 260 annotated PDF sequences were aligned on a phylogenetic tree, resulting in nine different phylogenetic groups. Long nodes compared to branches pointed to a low overall identity of peptide deformylases, which is in agreement with the knowledge that only few residues are essential for activity [7;21]. Despite this low sequence identity and the differences in secondary structure reported among PDFs, their overall 3D structure and catalytic properties are conserved. Exceptions are the mitochondrial PDF enzymes. Human and animal peptide deformylases contain mutations in the vicinity of the active site which are responsible for their low activity [25]. Mutations in regions that are highly conserved in bacterial PDF sequences have also been reported in *Trypanosomes* and *Archeae*, resulting in a low activity of these enzymes, which were found dispensable for growth [6;25]. We therefore focused on eubacterial PDFs.

In order to test the biocatalytic potential of different peptide deformylases, the genes from eight of these enzymes were cloned and overexpressed in *E. coli* (Chapter 3). Genes from *H. influenzae*, *P. aeruginosa* and *H. pylori*, with higher similarity to EcPDF, were best expressed. The eight different enzymes were purified using affinity column chromatography and all appeared to be catalytically active. Of the enzymes tested only the PDF from *B. stearothermophilus* and *H. influenzae* showed a better stability at high temperatures. Activity tests were also performed in presence of co-solvents and at different pH values but, unfortunately, the activities of the various PDFs were comparable with that of EcPDF, giving little hope to obtain peptide deformylases with higher stabilities through genome mining.

The purified peptide deformylases were also tested for their activity towards a range of racemic *N*-formylated amines and amino acid derivatives with the aim of exploring their substrate specificity. From the results obtained it was clear that the substrate specificities of the PDFs were very similar despite these enzymes belonging to different phylogenetic classes. All enzymes showed a high specificity for formylated peptides having methionine as first amino acid. The narrow substrate specificity is actually not surprising given the central role of peptide deformylase in ribosomal peptide synthesis and the conserved structure of their active site. In view of the results obtained, the chances to find a natural peptide deformylase with broader substrate specificity

seem very low, and therefore a protein engineering approach was explored to change enzyme properties.

The substrate specificity of the EcPDF was altered by simultaneous site-saturation mutagenesis at multiple positions. This approach is based on the detailed knowledge of the structure and mechanism of an enzyme and can be applied to peptide deformylase since several crystal structures of PDFs have been solved [10;12;19] and the substrate recognition and the selectivity were comprehensively studied [21]. Ragusa and co-workers reported efficient deformylation of peptides with methionine in position P_1 only. In line with this, *N*-formylated peptides with phenylalanine or leucine in the P_1 position gave an activity that was two orders of magnitude lower, and deformylation rates with formylated charged amino acids can be up to seven orders of magnitude lower [4;21].

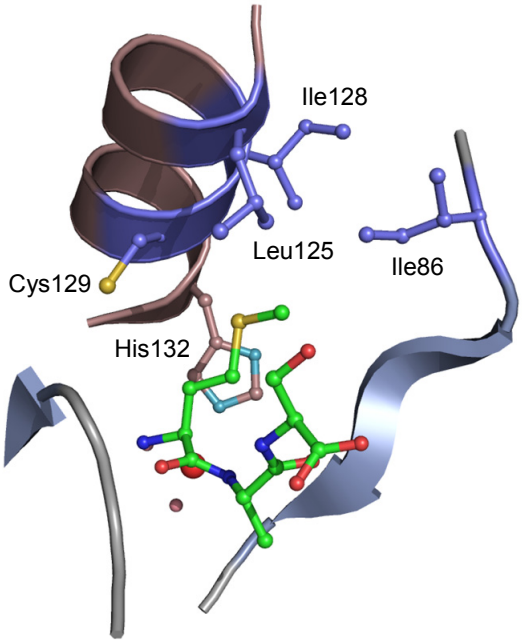


Figure 3. Active-site close-up of *E. coli* PDF with the positions of the residues targeted in the directed evolution library indicated. The ligand H-met-Ala-Ser-OH is shown in green.

By making use of the available structural information we were able to point at a set of amino acids delineating a hydrophobic pocket that lodges the methionine side chain in position P_1 of the nascent polypeptide. Of the seven amino acids composing this pocket only two residues were found to be highly conserved among PDFs from different organisms. These conserved residues are involved in the hydrogen bond

network that contributes to the correct 3D folding of the enzyme and in the binding of the metal ion responsible for catalysis, and thus they are crucial for enzyme function. The other residues of the pocket are much more variable among PDFs, although the selectivity of the enzymes remains identical. This information suggests that the overall shape and hydrophobic character of the pocket is responsible for the high specificity. Based on these data and on the crystal structure of EcPDF wild type, four amino acid residues in the hydrophobic pocket of $EcPDF_{Etag}$ (Ile-86, Leu-125, Ile-128 and Cys-129) were selected as promising targets for simultaneous saturation mutagenesis, generating a library of ~220,000 mutants (Chapter 5). By simultaneously saturating these sites, a library of high diversity targeting the hydrophobic S_1 position was created. At the same time, the combinatorial nature of the engineering approach chosen allows synergistic effects that can not be explored when mutating individual positions.

To be able to identify the improved variants in such a large mutant library, a prerequisite is to find a way to link the desired feature of a variant with the gene encoding it. This can be achieved by screening or selection methods. In Chapter 4 we investigated both approaches and describe the development of a selection method based on the complementation of *E. coli* leucine auxotrophy which was finally used for the library analysis. When large numbers of mutants need to be handled, a selection strategy that only enriches variants possessing the desired function, is preferable. Selection methods have the advantage of being less labour intensive than screenings since selection is based on a direct link between the desired property of the mutant and the survival capacity of the cell expressing this mutant [27;29]. In the case of peptide deformylase this link was created by making use of leucine-containing *N*-formylated dipeptides as sole source of leucine for the complementation of *E. coli* leucine auxotrophic strains. The design of a minimal medium that supports cell growth by expression of the enzyme variants of interest turned out to be quite difficult since several aspects needed to be taken into account. First the M9 minimal medium chosen needed to be supplemented with thiamine, amino acids and other trace elements in order to counterbalance the level of stress applied on the cells by the simultaneous growth and overexpression of peptide deformylase. Also the choice of the carbon source appeared crucial for obtaining good overproduction of PDF without negative effects on growth.

In the directed evolution studies we used an expression system based on the L-arabinose operon (pBAD vector). Glycerol is generally the preferred carbon source when this system is used since glucose inhibits protein production by catabolite repression [13;14;23]. However, when cells were grown in media containing glycerol, the growth-inhibiting effects of the induction were still present. Therefore, to further decrease the stress applied on the cells, glucose was used as carbon source [30]. In fact it is known that the *E. coli* growth rate is 2-fold higher in presence of glucose, while the total rate of gene transcription is higher when a carbon source of lower quality is used due to the activation of pathways linked to stress response in nutrient-poor environments [16;20]. During medium optimization the level of L-arabinose as inducer was increased to an optimal concentration for growth of *E. coli* in glucose containing medium and overexpression of peptide deformylase.

From the four-site randomization library, a total of about 400,000 CFU were plated on minimal medium containing *N*-formyl-Phe-Leu-OH as sole leucine source. Three mutants were found that grew faster on the selection medium, and they produced a PDF variant with a 2.5-fold increased activity as compared to $EcPDF_{Etag}$. All of them carried the same single mutation C138V, but encoded by 3 different codons reflecting the high quality of the mutant library. The same selection method allowed us to find four mutants that produced a PDF variant with a 2-fold increased activity towards *N*-formyl-Leu-Phe-NH_2, which was used as leucine source during selection. In this case the genes appeared to contain the mutation C138A.

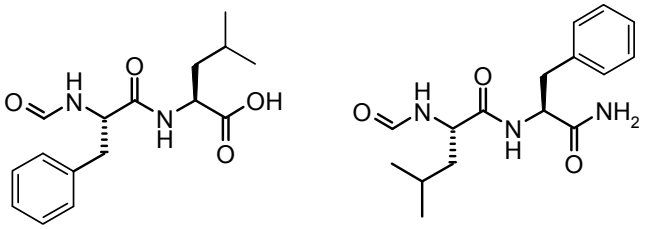


Figure 4. Substrates used for selection of PDF variants

The observation that the screening of this extensively mutated PDF library (4 positions randomized, 220,000 mutants) with two different selection substrates only gave mutants in which Cys-138 was replaced, triggered us to further explore the role of this cysteine in substrate recognition. By mutating Cys-138 in $EcPDF_{Etag}$ to all other 19 amino acids, seven mutants with remarkably different activities compared to wild-type were obtained. These results confirmed the key role of Cys-138 and the efficacy of the

selection method. At the same time the fact that very few mutations yielded active variants showed that the targeted residues form a delicate binding site that is very sensitive to deleterious effects of individual mutations. This makes it difficult to modify the substrate specificity of the enzyme while maintaining its activity. In view of the large sequence diversity of bacterial PDFs, this lack of resilience to mutations that alter the specificity of the enzyme is somewhat unexpected and suggests that individual PDFs have been well-optimized for selective methionine deformylation.

The development of peptide deformylase mutants with altered substrate specificity may require mutagenesis of Cys-138 in combination with a larger number of residues, which may possibly lead to more drastic redesign of the methionine side-chain binding pocket. It remains to be seen if this is possible without destroying the interactions required for catalysis. One way to prevent accumulation of negative mutations in such libraries may be computational analysis of enzyme stability and active site geometry, which may help to reduce library size and increase the frequency of positives. Further sequence and activity analysis of libraries such as those described in this thesis will help to identify deleterious mutations, which then can be omitted from future mutant collections. The beneficial mutations described here may be explored in combination with second shell mutations. Also a better understanding of the stubborn behaviour of *E. coli* PDF with respect to accepting mutations that provide it with an expanded substrate range may help the future construction of better variants.

References

1. **Albericio, F.** 2004. Developments in peptide and amide synthesis. Curr. Opin. Chem. Biol. 8:211-221.
2. **Albericio, F.** 2000. Orthogonal protecting groups for *N*(alpha)-amino and *C*-terminal carboxyl functions in solid-phase peptide synthesis. Biopolymers **55**:123-139.
3. **Becker, A., I. Schlichting, W. Kabsch, S. Schultz, and A. F. Wagner**. 1998. Structure of peptide deformylase and identification of the substrate binding site. J. Biol. Chem. **273**:11413-11416.
4. **Becker, A., I. Schlichting, W. Kabsch, D. Groche, S. Schultz, and A. F. V. Wagner**. 1998. Iron center, substrate recognition and mechanism of peptide deformylase. Nat. Struct. Mol. Biol. **5**:1053-1058.
5. **Ben Bassat, A., K. Bauer, S. Y. Chang, K. Myambo, A. Boosman, and S. Chang**. 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J. Bacteriol. **169**:751-757.

6. **Bouzadidi-Tiali, N., C. Giglione, Y. Bulliard, M. Pusnik, T. Meinnel, and A. Schneider**. 2007. Type 3 peptide deformylases are required for oxidative phosphorylation in Trypanosoma brucei. Mol. Biol. **65**:1218-1228.

7. **Dardel, F., S. Ragusa, C. Lazennec, S. Blanquet, and T. Meinnel**. 1998. Solution structure of nickel-peptide deformylase. J. Mol. Biol. **280**:501-513.

8. **Escobar-Alvarez, S., J. Gardner, A. Sheth, G. Manfredi, G. Yang, O. Ouerfelli, M. L. Heaney, and D. A. Scheinberg**. 2010. Inhibition of human peptide deformylase disrupts mitochondrial function. Mol. Cell. Biol. **30**:5099-5109.

9. **Escobar-Alvarez, S., Y. Goldgur, G. Yang, O. Ouerfelli, Y. Li, and D. A. Scheinberg**. 2009. Structure and activity of human mitochondrial peptide deformylase, a novel cancer target. J. Mol. Biol. **387**:1211-1228

10. **Fieulaine, S., C. Juillan-Binard, A. Serero, F. Dardel, C. Giglione, T. Meinnel, and J. L. Ferrer**. 2005. The crystal structure of mitochondrial (type 1A) peptide deformylase provides clear guidelines for the design of inhibitors specific for the bacterial forms. J. Biol. Chem. **280**:42315-42324.

11. **Giglione, C**. 2000. Identification of eukaryotic peptide deformylases reveals universitality of N-terminal protein processing mechanisms. EMBO J. **19**:5916-5929.

12. **Guilloteau, J. P., M. Mathieu, C. Giglione, V. Blanc, A. Dupuy, M. Chevrier, P. Gil, A. Famechon, T. Meinnel, and V. Mikol**. 2002. The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: a platform for the structure-based design of antibacterial agents. J. Mol. Biol. **320**:951-962.

13. **Hirsh, J. and R. Schleif**. 1973. In vivo experiments on the mechanism of action of -arabinose C gene activator and lactose repressor. J. Mol. Biol. **80**:433-444.

14. **Hodgson, D.** 1982. Glucose repression of carbon source uptake and metabolism in *Streptomyces coelicolor* A3(2) and its perturbation in mutants resistant to 2-deoxyglucose. J. Gen. Microbiol. **128**:2417-2430.

15. **Leeds, J. A. and C. R. Dean**. 2006. Peptide deformylase as an antibacterial target: a critical assessment. Curr. Opin. Pharmac. **6**:445-452.

16. **Liu, M., T. Durfee, J. E. Cabrera, K. Zhao, D. J. Jin, and F. R. Blattner**. 2005. Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. J. Biol. Chem. **280**:15921-15927.

17. **Lombard, C., J. Saulnier, and J. M. Wallach**. 2005. Recent trends in protease-catalyzed peptide synthesis. Protein Peptide Lett. **12**:621-629.

18. **Muramatsu, I., M. Murakami, T. Yoneda, and A. Hagitani**. 1964. The formylation of amino acids with acetic formic anhydride. Bull. Chem. Soc. Jpn. **38**:244-246.

19. **Nam, K. H., K. H. Kim, E. E. K. Kim, and K. Y. Hwang**. 2009. Crystal structure of an *Ef*PDF complex with Met-Ala-Ser based on crystallographic packing. Biochem. Biophys. Res. Commun. **381**:630-633.

20. **Paliy, O. and T. Gunasekera**. 2007. Growth of *E. coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents. Appl. Microbiol. Biotechnol. **73**:1169-1172.

21. **Ragusa, S., P. Mouchet, C. Lazennec, V. Dive, and T. Meinnel**. 1999. Substrate recognition and selectivity of peptide deformylase. similarities and differences with metzincins and thermolysin. J. Mol. Biol. **289**:1445-1457.

22. **Rajagopalan, P. T. R., A. Datta, and D. Pei**. 1997. Purification, characterization, and inhibition of peptide deformylase from *Escherichia coli*. Biochemistry **36**:13910-13918.

23. **Schleif, R.** 2000. Regulation of the arabinose operon of *Escherichia coli*. Trends Genetics **16**:559-565.

24. **Serero, A., C. Giglione, and T. Meinnel**. 2001. Distinctive features of the two classes of eukaryotic peptide deformylases. J. Mol. Biol. **314**:695-708.

25. **Serero, A., C. Giglione, A. Sardini, J. Martinez-Sanz, and T. Meinnel**. 2003. An unusual peptide deformylase features in the human mitochondrial *N*-terminal methionine excision pathway. J. Biol. Chem. **278**:52953-52963.

26. **Sinisterra, J. V. and A. R. Alcantara**. 1993. Synthesis of peptides catalysed by enzymes: a practical overview. J. Mol. Catal. **84**:327-364.

27. **Sonke, T., L. Duchateau, D. Schipper, G.J. Euverink, S. van der Wal, H. Henderickx, R. Bezemer, and A. Vollebregt**. 2009. Enzyme assays: high-throughput screening, genetic selection and fingerprinting. Wiley-VCH.

28. **Sonke, T., B. Kaptein, A. F. V. Wagner, P. J. L. M. Quaedflieg, S. Schultz, S. Ernste, A. Schepers, J. H. M. Mommers, and Q. B. Broxterman**. 2004. Peptide deformylase as biocatalyst for the synthesis of enantiomerically pure amino acid derivatives. J. Mol. Catal. B: Enzymatic **29**:265-277.

29. **Taylor, S., P. Kast, and D. Hilvert**. 2001. Investigating and engineering enzymes by genetic selection. Angew. Chem. Int. Ed. Engl. **40**:3310-3335.

30. **Xu, Z., W. Shen, H. Chen, and P. Chen**. 2005. Effects of medium composition on the production of plasmid DNA vector potentially for human gene therapy. J. Zhejiang University Science **6B**:396-400.

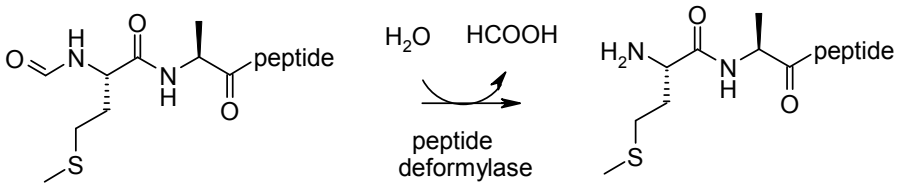
Samenvatting

Peptiden zijn polymere ketens die zijn samengesteld uit aminozuren die aan elkaar zijn gekoppeld via peptide-bindingen. Veel natuurlijke en synthetische peptiden hebben een biologische werking, bijvoorbeeld katalytische aktiviteit (enzymen), antibacteriële werking (antibiotica), of aktiviteit als hormoon (b.v. insuline). Vanwege hun effecten op tal van fysiologische processen kunnen peptiden medisch toegepast worden, zoals bij de regeling van de bloedruk of de bloedsuikerspiegel. Ook zijn er peptiden die de vruchtbaarheid beinvloeden. Daarnaast zijn peptiden van toenemend belang in voeding, vooral vanwege smaak- en gezondheidseffecten. In de meeste gevallen gat het om kleine peptiden, bv. van 2-50 aminozuren in lengte. Peptiden die veel langer zijn worden meestal eiwitten genoemd.

De toenemende interesse in het gebruik van peptiden in medische produkten en voedingsmiddelen leidt tot een groeiende behoefte aan methoden waarmee deze stoffen op een goedkope en veilige wijze gesynthetiseerd kunnen worden. Ook is er behoefte aan betere isolatie- en analyse-technieken voor peptiden. Er zijn drie principieel verschillende methoden voor de synthese van kleine peptiden: chemische synthese, produktie door fermentatie, en chemo-enzymatische synthese. Fermentatie is vooraal geschikt voor grotere peptiden en eiwitten. Korte peptiden kunnen redelijk efficiënt geproduceerd worden door chemische synthese. Chemo-enymatische peptide-synthese is vooral van aantrekkelijk voor de produktie van middelgrote peptiden, b.v. van 5-20 aminozuren. Daarbij worden enzymen gebruikt voor het aan elkaar koppelen van individuele aminozuren of van kleine peptiden die eerst chemisch zijn gesynthetiseerd. De enzymatische vorming van de peptide-bindingen vereist gewoonlijk dat de te koppelen carboxyl-groepen vooraf geaktiveerd worden (b.v. als een ester). Daarnaast is het nodig carboxyl-groepen en amine-groepen die niet moeten reageren te beschermen met een zgn. beschermgroep. De meeste beschermgroepen zijn duur in gebruik, en hebben vaak nadelen die industrieel gebruik uitsluiten.

Een potentieel aantrekkelijke beschermgroep voor de vrije N-terminale amine van een peptide of aminozuur is de formyl-groep (For). Deze is aantrekkelijk omdat de formyl-groep (-C(O)H) goedkoop is, en in principe enzymatisch verwijderd kan worden. Dat laatste is van belang omdat milde condities voorkomen dat het peptide racemiseert of anderszins beschadigd raakt tijdens de synthese. Dit proefschrift beschrijft een onderzoek naar de enzymatische verwijdering van de formyl-groep met het enzym peptide deformylase.

Het belang van peptiden, verschillende aspecten van de produktie van peptiden, en de eigenschappen van peptide deformylase zijn beschreven in **Hoofdstuk 1**. Peptide deformylase (PDF) is in staat een formyl-groep af te splitsen zonder de interne peptide-binding te splitsen. Deze selectiviteit vindt zijn oorsprong in de biologische functie van het enzym: bacteriële ribosomen (en eukaryote organellen) synthetiseren eiwitten die als N-terminus een *N*-formylmethionine hebben. Na de synthese van het peptide wordt de *N*-formylgroep afgesplitst door peptide deformylase, dat dus specifiek is voor peptiden die beginnen met (*N*-formyl)methionine.



Peptide deformylase reactie met een peptide dat begint met de sequentie For-Met-Ala.

Toepassing van peptide deformylase bij industriele produktie van peptiden vereist dat het enzym een bredere specificiteit heeft, dus niet alleen een *N*-formyl-groep kan afsplitsen van een peptide dat begint met *N*-formylmethionine, maar ook met peptiden die een ander geformyleerd N-terminaal aminozuur hebben. Ook is het nodig PDF op een goedkope manier in grote hoeveelheden te producere en te zuiveren. Daarbij is vooral van belang dat industriële preparaten geen peptidasen als verontreiniging bevatten die de te synthetiseren polypeptides zouden kunnen afbreken. Wanneer het gastheer-organisme waarin enzym-produktie plaatsvindt peptidasen bevat, moeten deze tijdens de zuivering dus efficiënt gescheiden worden van het peptide deformylase.

De ontwikkeling van een methode voor de snelle zuivering van peptide deformylase van de bacterie *E. coli* is beschreven in **Hoofdstuk 2** van dit proefschrift. Om de kosten laag te houden is vooral gekeken naar ultrafiltratie (UF) en anionen-wisseling-chromatografie (IEC). PDF is een klein eiwit (168 aminozuren). Om zuivering met ultrafiltratie te verbeteren is een nog kortere (kleinere) variant van het PDF gemaakt. Deze variant, die een C-terminale alpha-helix mist, leverde wel een betere zuivering en scheiding van amidasen op dan met het oorspronkelijke enzyme, maar de opbrengst bij zuivering was nog steeds slechts 21%, en de scheiding van peptidasen en PDF was nog onvoldoende. Daarom is een ook een andere aanpak gevolgd: op gen-

niveau is er een polyglutamaat-tag aangebracht in een flexibel gedeelte van het PDF. De extra negatieve lading die hiermee geintroduceerd is, maakte het mogelijk een goede scheiding met IEC te bereiken. Zowel de opbrengst (61%) als de verwijdering van peptidasen (1300 maal minder dan in celvrij extract) waren beter dan met de UF methode.

Belangrijke eigenschappen voor industriële toepassing zijn stabiliteit en katalytische aktiviteit. Een manier om een PDF in handen te krijgen met een verhoogde stabiliteit is beschreven in **Hoofdstuk 3**. Als bron voor PDFs is gekozen voor microbiële genoomsequenties. Door in databases genen te zoeken die verwant zijn aan PDF van *E. coli* is een 8-tal varianten geidentificeerd, die vervolgens tot overexpreesie zijn gebracht in *E. coli* en gezuiverd m.b.v. affiniteitschromatografie. Na gedeeltelijke zuivering bleken de enzymen te verschillen wat betreft temperatuur-stabiliteit, maar de nieuwe varianten waren nauwelijks beter wat betreft resistentie tegen organische oplosmiddelen en substraatbereik. Resistentie tegen oplosmiddelen is van belang omdat veel peptiden slecht oplosbaar zijn in water; toevoeging van een oplosmiddel kan de oplosbaarheid verhogen. De specificiteit van de gezuiverde enzymen is getest en het bleek dat alle PDFs een sterke voorkeur hadden voor *N*-formyl-peptiden met methionine als eerste aminozuur. Dit is niet verrassend omdat alle ribosomaal-gesynthetiseerde eiwitten, die de natuurlijke substraten zijn, ook altijd aan de N-terminus een *N*-geformyleerde metionine bezitten.

Een andere aanpak voor het in handen krijgen van een PDF met een breder substraatbereik is protein engineering. Dit kan op een gerichte manier, door mutaties te maken op grond van de bekende kristalstruktuur van PDF, of met een random aanpak (directed evolution). In dit proefschrift is een hybride aanpak gevolgd: het random maken van een beperkt aantal posities rond de active site van PDF. Zo'n aanpak vereist een selectie- of screening-methode waarmee snel de gewenste varianten kunnen worden geidentificeerd in een collectie van een zeer groot aantal gemuteerde enzymen. In **Hoofdstuk 4** beschrijft het proefschrift de ontwikkeling van dergelijke selectie-en screening-technieken. De beste selectie-methode is gebaseerd op een complementatie van de leucine auxotrofie van een *E. coli* gastheer. Door aan de *E. coli* cellen, die voor groei afhankelijk zijn van een bron van leucine, een geformyleerd dipeptide aan te bieden waaruit leucine alleen kan worden vrijgemaakt als de formyl groep eerst enzymatisch wordt afgesplitst, is een selectie mogelijk van cellen die een PDF met

gemodificeerde selectiviteit bezitten. Toepassing van deze methode vereiste aanpassing van het groei-medium zodanig dat stress door overexpressie van PDF werd onderdrukt. Ook een aantal andere methoden voor de identificatie van PDF met veranderde selectiviteit is getest.

Het gebruik van dit selectiesysteem voor de identificatie van mutanten van PDF is beschreven in **Hoofdstuk 5**. Doel was mutanten te vinden met een bredere specificiteit zodat *N*-formyl peptiden die niet beginnen met een (geformyleerde) methionine ook gedeformyleerd kunnen worden. Een library van PDF varianten is gemaakt door vier posities simultaan random te maken, hetgeen een zeer grote library leverde, waarin theoretisch 160,000 varianten aanwezig kunnen zijn. De gekozen posities vormen een hydrofobe pocket in de active site waarin de zijgroep van methionine past. De aminozuren op deze posities zijn niet geconserveerd in PDFs, hetgeen suggereert dat ze gemuteerd kunnen worden zonder de aktiviteit van het PDF te vernietigen. De library zal dus varianten bevatten met grote verschillen in de geometrie van deze methionine-bindende pocket.

De library were gescreened met de genoemde groei-selectie strategie. Verschillende clones met een veranderd substraatbereik zijn gevonden. Na sequentie-analyse bleek dat alle mutanten op positie Cys138 waren gemuteerd. De selectiviteit van deze mutant-enzymen was inderdaad veranderd, hetgeen aantoonde dat de selectiestrategie werkte. De eigenschappen van de mutanten lieten verder zien dat deze postie (Cys138) uiterst belangrijk is voor substraat-herkenning. Helaas leverde deze aanpak tot nu toe slechts een beperkt aantal aktieve varianten op. De zijgroep-pocket moet dus erg gevoelig zijn voor mutaties. Mogelijk is een groter aantal (compenserende) mutaties rond de methionine-pocket nodig om stabiele varianten te produceren met meer ingrijpende veranderingen in substraabereik.

.

Acknowledgements

This work would not have been possible without the help of many people that in one or another way contributed to my research. It is a pleasure for me to acknowledge them for their support, guidance and patience.

In the first place I would like to thank my promoter and co-promoter. Theo, thank you for choosing me as your PhD student. You have all my gratitude for the supervision, advice and encouragement during the years I spent at DSM. You always believed we could achieve great results with this project, to great results and you taught me, among many other things, how to be positive and passionate about my work (even on Monday mornings before your coffee in the Socio). Dick, I would like to thank you for the invaluable stimulating discussions and enthusiastic support you always gave me even at such long distance. I learned a lot from you, in particular accuracy in scientific writing.

I am also very grateful to Peter Quaedflieg for his scientific advice and knowledge and many insightful discussions and suggestions.

I would like to thank you the members of my reading committee Prof. dr. L. Dijkhuizen, Prof. dr. R. Bovenberg and Prof. dr. U. Hanefeld, for their kindness in reading and approving my work.

I gratefully acknowledge Dr. Volker Wagner, his input and deep knowledge about PDF has been a source of inspiration for this work.

Silvana, I will always be grateful for your support. Part of this thesis would not have been possible without your chemistry lessons, your enthusiasm and faith in the potential of “my” enzyme. Grazie mille per l’aiuto e l’amicizia che mi hai dato. John, I would like to thank you for your valuable input and knowledge on the PDF variants.

A special thank you goes to Linda and Monika, my office mates. I had such a good time in that little box. Your support and friendship have been of great help during the Dutch rainy days.

And of course Martin, always calm, inspiring and motivating. Thank you for always finding the time to answer to my questions. I really learned a lot from you.

Wilco, it’s strange how I still have good memories of the fermentation room, with its terrible smell and the noise of the sonicator, but I had some good laughs there with you. Theo, I hope I still have the title of best Italian girl behind the closed doors and yes, I do eat banana jam to keep healthy. Thanks to both of you.

And of course all the colleagues “behind the klapdeuren”, Elise, Elly, Natacha you made me feel welcome and home. Harry, I hope you did not miss too much my philosophy

classes in the lab. Claudia, grazie per le lezioni di chimica e le chiacchierate nei momenti difficili. Ilse, it was great working with you. Your experience and professionalism have been so important for my work. I am thankful for the time you took to introduce me to the word of analysis and HPLC and for the effort you put in tolerating my messy desk and endless runs up and down the corridor. And you Lavinia, anche se alla fine questo caffè non l'abbiamo ancora bevuto, ti vorrei ringraziare per l'amicizia e il sostegno che mi hai sempre Saputo dare.

Thanks to all my colleagues in the former ASC&D: for my first DSM Classic, for the good time during the department weekends, for my first TLC and for bearing with my Italian loudness and hand gestures during the 4 years.

A special thank you goes to all my friends in Maastricht. I had the chance to meet many good people there. Luigi, Thank you for telling me "Don't go for a PhD, it's a sequence of painful moments" and for being there during all of them. Ursula, Jacob, I know I do not like games but I shared with you music, laughs, dinners and I hope we will keep this going for many more years. Maria, I would not have survived without your pet animals keeping me company when I started my PhD, thank you for welcoming me at your place, it did feel like home! Laurent, it's not the destination but the journey that matters. This was a long one but I learned so much on the way. For sure it was worth. Thank you for looking after me and for being a sincere friend. George, a great neighbor, a fun travel mate, always ready to help in each situation and most of all a superb windsurfing teacher.

Chiachun, thank you for the long talks, the great Chinese food and a quick introduction to style and art. Masa, Sayuri, you made me discover the word of *ce n'est pas elegant* and *sumimasen*, a great experience! Jim, thank you for your enigmatic humor (I know I am still far from getting it) and for trying to find the good intention behind my sometimes sharp sentences.

I would like to mention my two paranymphs. Thank you for being good friends. Natascha, yes, I made it! Thank you for believing in me during all these years, for your continuous support and for having allowed me and my mountains of papers in your living room so many months! Francesca, grazie per aver sopportato i miei sfoghi, per i suggerimenti che mi hai saputo dare e per aver condiviso con me un pizzico di filosofia e tanti momenti felici.

Ed infine la mia famiglia: mamma, papa, Andrea, grazie per aver accettato le mie scelte anche quando non sembravano logiche e per essermi stati vicini anche da lontano.

Claudia